Title: Exome chip meta-analysis fine maps causal variants and elucidates the genetic architecture of rare coding variants in smoking and alcohol use.

Running Title: Exome Meta-Analysis of Smoking and Alcohol

Keywords: Tobacco, Nicotine, Alcohol, GWAS, Heritability, Behavioral Genetics

Number of words in abstract: 249
Number of words in main text: 3676
Number of Figures: 0
Number of Tables: 4
Number of Supplemental Materials: One Supplementary Note with eight supplementary tables and four supplementary figures.

Authors, in order with affiliation:

David M. Brazel*	Institute for Behavioral Genetics, University of Colorado Boulder Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder
Yu Jiang*	Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA
Jordan M. Hughey*	Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA
Valérie Turcot	Montreal Heart Institute, Montreal, Quebec, H1T 1C8, Canada Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada
Xiaowei Zhan	Department of Clinical Science, Center for Genetics of Host Defense, University of Texas Southwestern
Jian Gong	Public Health Sciences Division, Fred Hutchinson Cancer Research Center
Chiara Batini	Department of Health Sciences, University of Leicester
J. Dylan Weissenkampen	Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA
MengZhen Liu	Department of Psychology, University of Minnesota

CHD Exome+ Consortium[††]
Consortium for Genetics of Smoking Behaviour[††]

Daniel R. Barnes	Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
Sarah Bertelsen	Department of Neuroscience, Icahn School of Medicine at Mount Sinai
Yi-Ling Chou	Washington University
A. Mesut Erzurumluoglu	Department of Health Sciences, University of Leicester

This is the author's manuscript of the article published in final edited form as:
Brazel, D. M., Jiang, Y., Hughey, J. M., Turcot, V., Zhan, X., Gong, J., … Vrieze, S. (2018). Exome chip meta-analysis fine maps causal variants and elucidates the genetic architecture of rare coding variants in smoking and alcohol use. Biological Psychiatry. https://doi.org/10.1016/j.biopsych.2018.11.024

Jessica D. Faul	Survey Research Center, Institute for Social Research, University of Michigan
Jeff Haessler	Public Health Sciences Division, Fred Hutchinson Cancer Research Center
Anke R. Hammerschlag	Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam
Chris Hsu	University of Southern California
Manav Kapoor	Department of Neuroscience, Icahn School of Medicine at Mount Sinai
Dongbing Lai	Department of Medical and Molecular Genetics, Indiana University School of Medicine
Nhung Le	Department of Medical Microbiology, Immunology and Cell Biology, Southern Illinois University School of Medicine
Christiaan A de Leeuw	Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam
Anu Loukola	Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; Department of Public Health, University of Helsinki, Helsinki, Finland
Massimo Mangino	Department of Twin Research and Genetic Epidemiology, Kings College London, London SE1 7EH, UK; NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK
Carl A. Melbourne	Department of Health Sciences, University of Leicester
Giorgio Pistis	Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy.
Beenish Qaiser	Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; Department of Public Health, University of Helsinki, Helsinki, Finland
Rebecca Rohde	University of North Carolina, Chapel Hill
Yaming Shao	University of North Carolina, Chapel Hill
Heather Stringham	Department of Biostatistics, University of Michigan
Leah Wetherill	Department of Medical and Molecular Genetics, Indiana University School of Medicine
Wei Zhao	Department of Epidemiology, School of Public Health, University of Michigan
Arpana Agrawal	Department of Psychiatry, Washington University School of Medicine
Laura Bierut	Department of Psychiatry, Washington University School of Medicine
Chu Chen	Public Health Sciences Division, Fred Hutchinson Cancer Research Center

	Department of Epidemiology and Department of Otolaryngology; Head and Neck Surgery, University of Washington, Seattle, WA
Charles B. Eaton	Department of Family Medicine, Brown University, Providence, RI
Alison Goate	Department of Neuroscience, Icahn School of Medicine at Mount Sinai
Christopher Haiman	Department of Preventative Medicine, Keck School of Medicine, University of Southern California
Andrew Heath	Department of Psychiatry, Washington University
William G. Iacono	Department of Psychology, University of Minnesota
Nicholas G. Martin	Queensland Institute for Medical Research
Tinca J. Polderman	Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam
Alex Reiner	Public Health Sciences Division, Fred Hutchinson Cancer Research Center Department of Epidemiology, University of Washington, Seattle, WA
John Rice	Departments of Psychiatry and Mathematics, Washington University St. Louis
David Schlessinger	National Institute on Aging, National Institutes of Health
H Steven Scholte	Department of Psychology, University of Amsterdam & Amsterdam Brain and Cognition, University of Amsterdam
Jennifer A. Smith	Department of Epidemiology, School of Public Health, University of Michigan
Jean-Claude Tardif	Montreal Heart Institute, Montreal, Quebec, H1T 1C8, Canada Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada
Hilary A. Tindle	Department of Medicine, Vanderbilt University, Nashville, TN
Andries R van der Leij	Department of Psychology, University of Amsterdam & Amsterdam Brain and Cognition, University of Amsterdam
Michael Boehnke	Department of Biostatistics, School of Public Health, University of MIchigan
Jenny Chang-Claude	Division of Cancer Epidemiology, German Cancer Research Center
Francesco Cucca	Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy.
Sean P. David	Department of Medicine, Stanford University, Stanford, CA
Tatiana Foroud	Department of Medical and Molecular Genetics, Indiana University School of Medicine
Joanna M.M. Howson	Department of Public Health and Primary Care, University of Cambridge
Sharon L.R. Kardia	Department of Epidemiology, School of Public Health, University of Michigan
Charles Kooperberg	Public Health Sciences Division, Fred Hutchinson Cancer Research Center

Markku Laakso University of Eastern Finland, Finland
Guillaume Lettre Montreal Heart Institute, Montreal, Quebec, H1T 1C8, Canada
Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada
Pamela Madden Department of Psychiatry, Washington University
Matt McGue Department of Psychology, University of Minnesota
Kari North Department of Epidemiology, University of North Carolina, Chapel Hill
Danielle Posthuma Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam
Department of Clinical Genetics, VU University Medical Centre Amsterdam, Amsterdam Neuroscience
Timothy Spector Department of Genetic Epidemiology, Kings College, London
Daniel Stram Department of Preventative Medicine, Keck School of Medicine, University of Southern California
Martin D. Tobin Department of Health Sciences, University of Leicester
David R. Weir Survey Research Center, Institute for Social Research, University of Michigan
Jaakko Kaprio Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; Department of Public Health, University of Helsinki, Helsinki, Finland
Gonçalo R. Abecasis Regeneron Pharmaceuticals; Department of Biostatistics, School of Public Health, University of Michigan

Dajiang J. Liu† Institute of Personalized Medicine, Penn State College of Medicine
Scott Vrieze† Department of Psychology, University of Minnesota
*These authors contributed equally to the work.

†Address correspondence to Scott Vrieze (vrieze@umn.edu), University of Minnesota, 75 East River Road, Minneapolis, MN 55455; or Dajiang J. Liu (dajiang.liu@psu.edu), Penn State College of Medicine, HCAR 2020, Hershey, PA.

††See acknowledgments for a list of authors associated with the replication consortia.

Abstract:

Background: Smoking and alcohol use have been associated with common genetic variants in multiple loci. Rare variants within these loci hold promise in the identification of biological mechanisms in substance use. Exome arrays and genotype imputation can now efficiently genotype rare nonsynonymous and loss of function variants. Such variants are expected to have deleterious functional consequences, and contribute to disease risk.

Methods: We analyzed ~250,000 rare variants from 16 independent studies genotyped with exome arrays and augmented this dataset with imputed data from the UK Biobank. Associations

were tested for five phenotypes: cigarettes per day, pack years, smoking initiation, age of smoking initiation, and alcoholic drinks per week. We conducted stratified heritability analyses, single-variant tests, and gene-based burden tests of nonsynonymous/loss of function coding variants. We performed a novel fine mapping analysis to winnow the number of putative causal variants within associated loci.

Results: Meta-analytic sample sizes ranged from 152,348-433,216, depending on the phenotype. Rare coding variation explained 1.1-2.2% of phenotypic variance, reflecting 11%-18% of the total SNP heritability of these phenotypes. We identified 171 genome-wide associated loci across all phenotypes. Fine mapping identified putative causal variants with double base-pair resolution at 24 of these loci, and between 3 and 10 variants for 65 loci. 20 loci contained rare coding variants in the 95% credible intervals.

Conclusions: Rare coding variation significantly contributes to the heritability of smoking and alcohol use. Fine mapping GWAS loci identifies specific variants contributing to the biological etiology of substance use behavior.

Introduction

Tobacco and alcohol use together account for more morbidity and mortality in Western society than any other single risk factor or health condition(1). These preventable and modifiable behaviors are heritable(2), but previous human and model organism research, including genome-wide association studies of common variants, have resulted in few associated genetic variants, which most prominently feature genes involved in alcohol/nicotine metabolism and nicotinic receptors(3-7).

Advances in sequencing, genotyping, and genotype imputation now allow cost effective investigation of rare and low frequency variants. Compared to common variants (minor allele frequency [MAF] > 1%) most commonly used in genome-wide association studies (GWAS), rare variants have greater potential to elucidate biological mechanisms of complex traits, including substance use and addiction(8, 9). In particular, nonsynonymous and loss of function (LoF) coding variants, which result in the loss of normal function of a protein, may have greater phenotypic impact and more direct mechanistic interpretation than other variants that do not have obvious biological consequences(10, 11).

No large-scale genome- or exome-wide study of rare variation has been conducted to date. The vast majority of existing addiction-related rare variant studies have used targeted sequencing of putative addiction-associated loci to discover and test for association in relatively small samples. Existing research has led to intriguing leads, including rare variant associations in loci that span nicotinic receptor gene clusters(12-21) and alcohol metabolism genes(22-24) for nicotine and alcohol dependence, respectively. This strategy has also produced rare variant associations in novel loci. In one case, gene-level association tests were used to find an association with rare variants in *SERINC2*(24). In another case, a burden test across *PTP4A1*, *PHF3*, and *EYS* showed association with alcohol dependence(25). Unfortunately, these genes are not obviously involved in etiological processes related to addiction, and replications have not been reported to date.

Previous studies have also attempted to leverage information about predicted functional consequences of rare mutations to improve association analyses. One study of nicotine dependence found significant rare single-variant associations in *CHRNB4*, but only when variants were weighted by their predicted effect on the cellular response to nicotine and acetylcholine(26). Such positive findings could benefit from replication, which has not always been straightforward. For example, all rare variant associations in addiction are, to our knowledge, candidate gene analyses with type I error thresholds based only on the number of tests within that region. Historically, such analyses have produced overly optimistic estimates of the number of associated loci(27). Genome-wide analyses with more conservative type I error thresholds have reported null rare variant findings across an array of phenotypes relevant to addiction(28-30). Precisely because genome-wide analyses are conducted on many variants across the genome, they are in principle able to discover novel rare variant associations within new or known loci. One way to improve power in genome-wide analyses is through genetic association meta-analysis, which entails the aggregation of results across many studies to achieve large sample sizes.

Here, we attempted to expand on these previous discoveries by conducting the largest meta-analytic investigation of exonic rare variants to date. We conducted an exome-wide association meta-analysis of nicotine and alcohol use across 16 studies genotyped on the exome array, which genotypes low-frequency nonsynonymous and putative loss of function exonic variants. We combined these data with the UK Biobank, which includes approximately 400,000 individuals of European ancestry with genotype imputation to the Haplotype Reference Consortium(31) imputation reference panel and relevant smoking/drinking phenotypes. Sample sizes for well-imputed variants were thus enlarged and the availability of noncoding variants from UK Biobank enabled comprehensive analysis of genetic architecture(32) and fine mapping(33).

We conducted single variant and gene-based tests of association with five smoking and drinking phenotypes. We applied a novel fine mapping analysis to prioritize causal variants using statistical and functional information. We also evaluated the contribution of rare exonic variants to the heritability of these phenotypes. Family studies, as well as studies of the aggregate effects of common variants, have found both alcohol use and tobacco use to be heritable behaviors(30, 34-38). Research on the aggregate contribution of rare variants, however, has been scarce, with previous work on related phenotypes in smaller samples failing to detect aggregate effects for smoking and alcohol consumption(28). We used meta-analytic summary statistics to quantify the contribution to heritability of variants in various functional categories and frequency bins.

Methods and Materials

Seventeen studies contributed summary statistics for meta-analysis. These studies, their sample sizes, and available phenotypes are listed in the online supplement (**Tables S1 and S2**). We augmented our sixteen exome chip cohorts with the UK Biobank, where imputation to the Haplotype Reference Consortium panel was used in lieu of an exome chip array. All individuals were of European ancestry, as determined by genetic principal components.

Phenotypes

Phenotypes were selected to represent multiple stages of smoking. These included initiation, heaviness of use among smokers, and a measure of total lifetime exposure to tobacco. For alcohol use only a measure of amount of alcohol use was systematically available across studies. The selected phenotypes are relevant to prior GWAS of smoking and alcohol use; are commonly available in psychological, medical, and epidemiological data sets; and are known to be correlated with measures of substance dependence(4, 39-41).

1. Cigarettes per day (CigDay). The average number of cigarettes smoked in a day among current and former smokers. Studies with binned responses used their existing bins. Studies that recorded an integer value binned responses into one of four categories: 1=1-10, 2=11-20,

3=21-30, 4=31 or more. Anyone reporting 0 cigarettes per day was coded as missing. This phenotype is a component of commonly used measures of nicotine dependence such as the Fagerstrom Test for Nicotine Dependence.

2. Pack Years (PckYr). Defined in the same way as cigarettes per day but not necessarily binned, divided by 20 (cigarettes in a pack), and multiplied by number of years smoking. This yielded a measure of total overall exposure to tobacco and is relevant to disease outcomes for which smoking is a risk factor, such as cancer and chronic obstructive pulmonary disease risk.

3. Age of Initiation of Smoking (AgeSmk). A measure of early cigarette use. Defined as the age at which a participant first started smoking regularly.

4. Smoking Initiation (SmkInit). A binary variable of whether the individual had ever been a regular smoker (1) or not (0), and often defined as having smoked at least 100 cigarettes during one's lifetime.

5. Drinks per week (DrnkWk). A measure of drinking frequency/quantity. The average number of drinks per week in current or former drinkers.

Genotypes

Fourteen of the seventeen studies were genotyped with the Illumina HumanExome BeadChip, which contains ~250,000 low-frequency nonsynonymous variants, variants from the GWAS catalog, and a small number of variants selected for other purposes. Two studies were genotyped on the Illumina Human Core Exome, which includes an additional ~250,000 tag SNPs. The remaining study, the UK Biobank, was imputed using Haplotype Reference Consortium panel(31, 42), as well as the reference panel by UK 10K and 1000 Genomes Project. An integrated callset was released by the UK Biobank team(42). Our UK Biobank genetic association analyses were conducted based on the integrated callset with additional quality control.

Generation of Summary Association Statistics

Seventeen independent studies (see **Table S1**) with smoking and drinking phenotypes were included in the discovery phase. Individual studies conducted association analysis accounting for age, sex, any study-specific covariates, and ancestry principal components (see **Table S2** for genomic controls), and submitted summary statistics for meta-analysis. For studies with related individuals (**see Table S1**), relatedness was accounted for in linear mixed models using empirically estimated kinships from common SNPs(43). Residuals were inverse-normalized to help ensure well-behaved test statistics for rare variant tests.

Quality control of per-study summary statistics included evaluation and correction of strand flips and allele flips through systematic comparison of alleles and allele frequencies against the reference datasets ExAC v2.0, 1000 Genomes Phase 3, and dbSNP. Variants with call rates<0.9, or Hardy Weinberg $p<1x10^{-7}$ were also removed. The latter filter was meant to avoid findings that could not be more broadly replicated across the 17 studies.

Meta-analysis

Association testing was done in stages. First, we conducted genome-wide association meta-analysis. Variants with p-values less than the genome-wide significance threshold of 5×10^{-8} were deemed statistically significant. Loci were defined as 1 million basepair windows surrounding a "sentinel" (most significant) variant in the locus. Overlapping or adjacent loci were combined into a single locus. Conditional analysis and fine mapping was then performed within each locus. We attempted to replicate one very rare variant (rs36015615 in *STARD3* associated with CigDay; see results and **Table 1**) that was available in two other exome chip consortia. These were the CHD Exome+ Consortium (N=17,789) and the Consortium for Genetics of Smoking Behaviour (N=28,583). Both consortia defined their phenotypes, including cigarettes per day similarly, as the usual number of cigarettes smoked in a day corrected for sex, age, principal components (and/or genetic relatedness, as appropriate), and inverse-normalized prior to association analysis.

We also conducted gene-level association tests grouping nonsynonymous, stop gain, stop loss and splice variants within each gene, using rareMETALS version 6.0(44). Variant annotation was conducted using SEQMINER with RefSeq 1.9(45). Two complementary gene-level association tests were performed: the sequence kernel association test (SKAT; 46, 47) with a MAF cutoff of 1% and a simple burden test(48) that summed the number of rare alleles within a given gene, again with a maximum MAF=1%. We chose variants with MAF≤1% as we were interested in the contribution of variants with a frequency lower than that which has been reliably imputed and tested in past GWAS meta-analyses. We considered a gene association to be significant if the p-value surpassed a Bonferroni correction for the number of genes tested for a given phenotype and test, assuming approximately 20,000 genes in the genome (.05/20,000 = 2.5×10^{-6}).

We performed iterative conditional analysis using a partial correlation based score (PCBS) statistic(49), which can perform proper conditional analysis for meta-analysis that combines datasets measured using different arrays. PCBS takes GWAS meta-analysis summary statistics and LD estimated from the Haplotype Reference Consortium panel as input.

As a key step to evaluate the contribution of variants within a genome-wide significant locus(33), we used our PCBS framework to apply two complementary fine mapping techniques to identify putatively causal genetic variants. The first technique was a Bayesian approach described previously(50) that estimates the posterior probability of association based upon the statistical strength of the association for variants in each locus. We also applied a version of fgwas(51) modified to work within PCBS, which assumes that variants in different functional categories have potentially different prior probability of association. For loci with a single association signal based, effect sizes and variance from single-SNP analyses were used. If a locus contained multiple signals, we used effect sizes and variance from conditional analysis adjusting for all other index variants in this region.

Finally, we attempted to replicate previous rare variant associations referenced in the introduction and listed in **Table S4.** We attempted replication in our phenotypes for any single variant when that variant was directly genotyped or imputed. We applied a liberal threshold that corrected only for the number of tests conducted for this replication exercise (.05/46=.001).

Genetic Architecture

We performed heritability and genetic correlation analyses using LD score regression(52). The method calculates LD scores from the Haplotype Reference Consortium and the estimation of heritability with these LD scores then follows established methods(53, 54). Heritability was estimated for each trait and partitioned by annotation category and frequency bins. First, we annotated variants on the exome chip based upon gene definitions in RefSeq 1.9, using SEQMINER version 6.0(55). A variant is classified as coding if it belongs to either one of the following categories: nonsynonymous, stop gain, stop loss, and splice. Seven functional categories were considered in the model, including intergenic, intron, common coding (MAF>0.01), rare coding (MAF<0.01), synonymous, and 3'/5' untranslated regions. We fitted the baseline model with seven categories, and estimated phenotypic variance explained by each category.

Results

GWAS analyses behaved well, with genomic control values for the GWAS across exome chip and UK Biobank imputed variants between 1.05 and 1.3. The intercept for LD Score regression ranged between .99 and 1.1, indicating absent or minimal effects of population stratification. (Per-study genomic controls can be found in **Table S2**.) A total of 171 loci were identified under the genome-wide significance threshold ($p<5\times10^{-8}$), including 3, 11, 17, 93 and 47 loci for AgeSmk, CigDay, PckYr, SmkInit, and DrnkWk. A list of all sentinel variants within each locus is shown in **Table S5**. QQ plots and Manhattan plots are available in **Figures S1 and S2**. (Additional exploratory GWAS meta-analysis of individuals with significant African ancestry are provided in the Supplementary Note [including up to 8,974 individuals from three

studies]; see also **Table S3, Figure S3 and S4.**) The genome-wide significant association results included known loci associated with smoking and alcohol use phenotypes. These included associations between smoking phenotypes and variants within the *CHRNA5-CHRNA3-CHRNB4* nicotinic receptor cluster, nicotine metabolism gene *CYP2A6,* and a locus near dopamine receptor *DRD2*. We also replicated previous associations between nonsynonymous variant rs1229984 in *ADH1B* and DrnkWk. Only one very rare variant was associated with any of our five phenotypes. This was rs36015615 (MAF=.0002), a nonsynonymous variant in *STARD3*, associated with CigDay ($p=3.2\times10^{-8}$). This novel variant did not replicate in either of two replication consortium datasets, the CHD Exome+ Consortium (N=17,789, Beta=-.01, *p*=.94) or the Consortium for Genetics of Smoking Behaviour (N=28,583, Beta=.056, *p*=.84). Based upon the estimated genetic effects in the discovery sample ($\beta = 1.2$), the power for replication is >99%. However, if we assume the observed effect sizes in the replication datasets are correct, there is 5% power for replication based upon this estimated effect. The pattern of results may be due to winner's curse, or the discovered variant may be a false positive finding. Additional studies are required to narrow the possible interpretations.

The fine mapping analysis of all 171 GWAS loci pinpointed putatively causal variants with high resolution in some cases. The 95% credible interval for 34% of the loci had <10 SNPs and 24 loci had double basepair resolution, including several instances where the sole putative causal variant was nonsynonymous and of lower frequency, although in only one case with MAF<1%. The resolution increased somewhat when functional information was used to inform the prior, with double base-pair resolution at 32 loci, and 44% of loci having <10 SNPs in the 95% credible interval. **Table 1** includes all nonsynonymous or loss of function variants within the genome-wide significant loci that had a posterior probability of association greater than .80 from at least one of the fine mapping methods. Additional results from the fine mapping analysis are available in **Tables S6 and S7**. Several known functional variants were identified through this method, including: rs16969968(56), a nonsynonymous variant in nicotinic receptor gene

CHRNA5 associated with CigDay (PPA=.92 and .84 from the fine mapping analysis with, and without, functional priors, respectively); rs1229984(57), a nonsynonymous variant in alcohol metabolism gene *ADH1B* associated with DrnkWk (PPA=1.0 and 1.0); and, although with somewhat weaker evidence, rs6265(58), a nonsynonymous variant in brain derived neurotrophic factor *BDNF* associated with SmkInit (MAF=.19; PPA=.83 and .32).

Novel variants in novel genes were also prioritized at high resolution. To take the most statistically compelling examples in **Table 1**, we found rs28929474, a low frequency nonsynonymous variant in *SERPINA1*, associated with DrnkWk (MAF=.02; PPA=1.0 and .95). When homozygous, the alternate T (allele frequency = .02; frequency of TT genotype under Hardy Weinberg = 4 in 10,000) allele is a leading cause of alpha-1 antitrypsin deficiency. Here, we find the same risk allele, the T allele, is associated with an approximately .05 standard deviation decrease in drinks per week. We also discovered rs35891966, a variant in *NAV2*, associated with SmkInit (MAF=.07; PPA = 1.0 and .98) at single base-pair resolution. *NAV2* is involved in neuronal development and previously shown to be differentially expressed between smokers and non-smokers, but not previously implicated in GWAS(59).

Results of gene-based tests are provided in **Table 2**. A novel gene, rho guanine nucleotide exchange factor 37 (*ARHGEF37*), was associated with Age of Initiation of Smoking ($p=1.9\times10^{-6}$). *ARHGEF37* has not been widely studied and its function in not well known. Another novel gene without an immediate biological interpretation, was HEAT Repeat Containing 5A (*HEATR5A*), associated with Smoking Initiation ($p=1.4\times10^{-8}$). We also discovered a significant gene-based association between known alcohol metabolism gene *ADH1C* and Drinks per Week ($p=1.4\times10^{-27}$ and $p=1.9\times10^{-40}$ from the burden and SKAT tests, respectively). Finally, even with relaxed p-value thresholds, we failed to replicate genes identified in previous rare variant association studies referenced in the introduction (**Table S4**), with the exception of *ADH1C* and *CHRNA5*, two loci long known to be associated with alcohol use and smoking, respectively.

The estimated total SNP heritability for AgeSmk, CigDay, PckYr, SmkInit, and DrnkWk was 6%, 9%, 10%, 14% and 16%. Significant phenotypic variance was explained by rare nonsynonymous variants for all traits, ranging from 1.0%-2.2% (**Table 3**). As a fraction of the SNP heritability, rare nonsynonymous variants accounted for 11%-18%. Results for all seven functional categories are listed in **Table S8**; appreciable heritability was accounted for by common and rare coding variants, and intergenic variants. Variants in the untranslated regions and intronic regions contributed less. Almost all pairs of phenotypes were genetically correlated (**Table 4, Panel A**), and the direction of the genetic correlations were in the expected direction. For instance, CigDay was positively correlated with DrnkWk (0.2 ± 0.09) , consistent with the observation that increased alcohol consumption is correlated with increased tobacco consumption. Age of initiation has a negative correlation with all other traits, which is consistent with the observation that an earlier age of smoking initiation is correlated with increased tobacco and alcohol consumption in adulthood. The patterns and magnitudes of correlation are highly similar when considering only rare nonsynonymous variants (**Table 4, Panel B**).

Discussion

With a maximum sample size ranging from 152,348 to 433,216, the present study is the largest study to date of low-frequency nonsynonymous and loss of function variants in smoking and alcohol use. Our meta-analytic study design combined studies genotyped on the exome array with imputed genotypes in the UK Biobank, allowed us to comprehensively evaluate the contribution of rare and low frequency variants to the etiology of tobacco and alcohol use. All told, we identified 171 genome-wide significant loci for the five phenotypes.

We showed that the rare variants (MAF≤1%) together explain 1.0% - 2.2% of the phenotypic variance for the five traits, amounting to 11-18% of the total SNP heritability. A number of putatively causal low frequency nonsynonymous variants in novel genes were identified through two complementary fine mapping techniques. These include a variant known

to affect alpha-1 antitrypsin deficiency in *SERPINA1*. The effect of the risk allele resulted in a decrease in drinks per week. One interpretation is that this variant leads to impaired liver function through alpha-1 antitrypsin deficiency which, in turn, reduces alcohol consumption. Interestingly, neither this particular variant or the locus surrounding it was associated with smoking phenotypes, even though alpha-1 antitrypsin deficiency also affects lung function over time. Other mechanisms by which *SERPINA1* exerts its effect on alcohol consumption are certainly possible. Another novel nonsynonymous variant was in neuron navigator 2 (*NAV2*), associated with smoking initiation. *NAV2* has not previously been associated with substance use or addiction. Given its suspected involvement in neuronal growth and migration, a putatively causal nonsynonymous variant is a strong candidate for functional follow up experiments. Other genes implicated in the fine mapping analysis have less direct interpretations (e.g., *HEATR5A*) and such results will benefit from replication and/or follow-up experiments. In general, fine mapping studies narrowed the credible set of likely causal variants to single or double base pair resolution for 24 loci (**Table S6**). Some loci were not amenable to fine mapping, with credible intervals containing thousands of SNPs in some cases. Given the cost in money and time of conducting functional experiments at the cellular or organismal level, fine mapping likely causal variants can be extremely useful in predicting functional consequences and prioritizing variants for further work.

Gene based tests identified a small number of associated genes, including an expected association with *ADH1C* and drinks per week. The other two associated genes, *ARHGEF37* and *HEATR5A*, do not lend themselves to ready biological interpretations.

We showed that rare coding variants available on the exome chip or imputable by the Haplotype Reference Consortium, with frequency <1%, explain significant proportions of phenotypic variance, and a substantial proportion of the total SNP heritability. The exome chip was designed to genotype coding variants uncovered in ~12,000 sequenced exomes. By design, it comprehensively ascertained high confidence rare nonsynonymous, splice, and stop

variants within those sequences and only sparsely genotypes other classes of variation, including common variants. The Haplotype Reference Consortium panel imputed data also have limited accuracy when the underlying genetic variants are rare. Therefore, our current investigation did not fully explore the genetic architecture of very rare variants (i.e. with MAF<0.1%). With the development of larger imputation reference panels, and the availability of large scale deep whole genome sequences (such as the Trans-Omics for Precision Medicine Study [TOPMed]), we expect to be able to conduct an even more comprehensive analysis of the genetic architecture for variants with ever lower frequencies. Ultimately, the discovery of low frequency with small effects will require even larger sample sizes. For example, for rare variant with MAF of .1% and effects of .2, .15, and 0.1 standard deviations on the phenotype, to identify associations at $\alpha = 5 \times 10^{-8}$ with 80% of power, sample sizes of 500,000 890,000 and 1,990,000 are required. While such numbers seemed astronomical just a few years ago, they will indeed be attainable in the next few years with the availability of large biobank datasets and ever improving imputation. Another limitation of the present study is the limited samples sizes from non-European ancestries, where only exploratory analyses were possible. Substantial improvements can be made to the resolution of fine mapping analysis by leveraging disparate LD information across samples with diverse ancestry(33). Future research will do well to include individuals of diverse ancestry.

Acknowledgements: Research reported in this article was supported by the National Institute on Drug Abuse and the National Human Genome Research Institute of the National Institutes of Health under award numbers R01DA037904 (SIV), R21DA040177 (DJL), R01HG008983 (DJL) R01GM126479 (DJL) and 5T3DA017637-13 (DMB), as well as funding sources listed in the Supplementary Note. JMH was supported by a NSF Graduate Research Fellowship. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Disclosures: There are no conflicts to disclose

CHD Exome+ consortium members

Praveen Surendran[1], Robin Young[1], Daniel R. Barnes[1], Sune Fallgaard Nielsen[2], Asif Rasheed[3], Maria Samuel[3], Wei Zhao[4], Jukka Kontto[5], Markus Perola[5,6,7], Muriel Caslake[8], Anton JM. de Craen[9], Stella Trompet[9,10], Maria Uria-Nickelsen[11], Anders Malarstig[12], Dermot F. Reily[13], Maarten Hoek[14], Thomas Vogt[14,15], J Wouter. Jukema[11,16], Naveed Sattar[17], Ian Ford[8], Chris J. Packard[8], Dewan S. Alam[18], Abdulla al Shafi. Majumder[19], Emanuele Di Angelantonio[1,20], Rajiv Chowdhury[1], Philippe Amouyel[21,22,23,24], Dominique Arveiler[25], Stefan Blankenberg[26,27], Jean Ferrières[28], Frank Kee[29], Kari Kuulasmaa[5], Martina Müller-Nurasyid[30,31,32], Giovanni Veronesi[33], Jarmo Virtamo[5], EPIC-CVD Consortium, Philippe Frossard[3], Børge Grønne Nordestgaard[2], Danish Saleheen[4,3,1], John Danesh[1,35,20], Adam S. Butterworth[1,20], Joanna MM. Howson[1]

Affiliations

1. Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
2. Department of Clinical Biochemistry Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark
3. Centre for Non-Communicable Diseases, Karachi, Pakistan
4. Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
5. Department of Health, National Institute for Health and Welfare, Helsinki, Finland
6. Institute of Molecular Medicine FIMM, University of Helsinki, Finland
7. Estonian Genome Center, University of Tartu, Tartu, Estonia
8. University of Glasgow, Glasgow, UK
9. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands
10. Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands
11. Development Management and Planning, Pfizer Worldwide Research and Development

12. Pfizer Worldwide Research and Development, Stockholm, Sweden
13. Genetics and Pharmacogenomics, Merck Research Laboratories, Boston, Massachusetts, USA.
14. Merck Research Laboratories, Kenilworth, New Jersey, USA
15. CHDI Management/CHDI Foundation, Princeton, New Jersey, USA
16. The Interuniversity Cardiology Institute of the Netherlands, Utrecht, The Netherlands
17. Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK
18. ICDDR, B; Mohakhali, Dhaka, Bangladesh
19. National Institute of Cardiovascular Diseases, Sher-e-Bangla Nagar, Dhaka, Bangladesh
20. The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK
21. University of Lille, Risk Factors and Molecular Determinants of aging-related diseases, Lille, France
22. Inserm, Lille, France
23. Centre Hospitalier Universitaire Lille, Public Health, Lille, France
24. Institute Pasteur de Lille, Lille, France
25. Department of Epidemiology and Public Health, EA 3430, University of Strasbourg, Strasbourg, France
26. Department of General and Interventional Cardiology, University Heart Center Hamburg, Germany

University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

Consortium for Genetics of Smoking Behaviour

A Mesut Erzurumluoglu [1], Victoria E Jackson [1], Carl A Melbourne [1], Tibor V Varga [2], Helen R Warren [3,4], Vinicius Tragante [5], Ioanna Tachmazidou [6], Sarah E Harris [7,8], Evangelos Evangelou [9,10], Jonathan Marten [11], Weihua Zhang [12,13,14,15], Elisabeth Altmaier [16], Jian'an Luan [17], Claudia Langenberg [17], Robert A Scott [17], Hanieh Yaghootkar [18], Kathleen Stirrups [19,20], Stavroula Kanoni [20,21], Eirini Marouli [20,21], Fredrik Karpe [22,23], Anna F Dominiczak [24], Peter Sever [25], Neil Poulter [26], Olov Rolandsson [27], Clemens Baumbach [16], Saima Afaq [12], John C Chambers [12,13,28], Jaspal S Kooner [29,13,30,31], Nicholas J Wareham [17], Frida Renström [2,32], Göran Hallmans [32], Riccardo E Marioni [7,8], Janie Corley [7,33], John M Starr [7,34], Niek Verweij [35,36], Rudolf A de Boer [35], Peter van der Meer [35], Ersin Yavas [37], Ilonca Vaartjes [38,39], Michiel L Bots [38,39], Folkert W Asselbergs [5,40], Hans J Grabe [41], Henry Völzke [42], Matthias Nauck [43], Stefan Weiss [44], Paul D P Pharoah [45,46], Alison M Dunning [46], Joe G Dennis [45], Deborah J Thompson [45], Kyriaki Michailidou [47,45], Douglas F Easton [45,46], Antonis C Antoniou [45], Jessica Tyrrell [18], Evelin Mihailov [48], Nilesh J Samani [49,50], Kaixin Zhou [51], Matthew J Neville [22,23], Andres Metspalu [48], Colin N A

Palmer [52], Ian P Hall [53], David P Strachan [54], Ian J Deary [7,33], Tim M Frayling [18], Caroline Hayward [11], Pim van der Harst [35,55], Eleftheria Zeggini [6], Understanding Society Scientific± Group , Patricia B Munroe [3,4], Jan-Håkan Jansson [56], Paul W Franks [2,57], Panos Deloukas [58,59,60], Mark J Caulfield [3,4], Louise V Wain [1], Martin D Tobin [1]

1. Department of Health Sciences, University of Leicester, Leicester, UK
2. Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre, Department of Clinical Sciences, Skåne University Hospital, Lund University, SE-214 28, Malmö, Sweden
3. Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK
4. NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, London, EC1M 6BQ, UK
5. Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, 3508GA Utrecht, the Netherlands
6. Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK
7. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK, EH8 9JZ
8. Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK, EH4 2XU
9. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK.
10. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece.
11. MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.
12. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK
13. Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex UB1 3HW, UK
14. Biocenter Oulu, University of Oulu, Finland.
15. Unit of Primary Care, Oulu University Hospital, Oulu, Finland
16. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany.
17. MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, CB2 0QQ, UK
18. Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom
19. Department of Haematology, University of Cambridge, Cambridge, UK, CB2 0PT
20. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK, EC1M 6BQ
21. Centre for Genomic Health, Queen Mary University of London, London EC1M 6BQ, UK
22. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford
23. Oxford National Institute for Health Research, Biomedical Research Centre, Churchill Hospital, Oxford, UK
24. Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK
25. National Heart and Lung Institute, Imperial College London, W2 1PG, UK
26. School of Public Health, Imperial College London, W2 1PG, UK
27. Department of Public Health & Clinical Medicine, Section for Family Medicine, Umeå universitet, SE-90185 Umeå, Sweden

28. School of Medicine and Pharmacology, The University of Western Australia, Crawley 6009, Australia
29. National Heart and Lung Institute, Imperial College London, London W12 0NN, UK
30. PathWest Laboratory Medicine of WA, Sir Charles Gairdner Hospital, Crawley WA 6009, Australia
31. School of Pathology and Laboratory Medicine, The University of Western Australia, Crawley WA 6009, Australia
32. Department of Biobank Research, Umeå University, SE-901 87, Umeå, Sweden
33. Psychology, University of Edinburgh, Edinburgh, UK, EH8 9JZ
34. Alzheimer Scotland Research Centre, University of Edinburgh, Edinburgh, UK, EH8 9JZ
35. University Medical Center Groningen, University of Groningen, Department of Cardiology, the Netherlands
36. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, 301 Binney Street, Cambridge, MA 02142, USA
37. Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, UK
38. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3508GA Utrecht, the Netherlands
39. Center for Circulatory Health, University Medical Center Utrecht, 3508GA Utrecht, the Netherlands
40. Durrer Center for Cardiovascular Research, Netherlands Heart Institute, 3501DG Utrecht, Netherlands
41. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, 17475 Greifswald, Germany
42. Institute for Community Medicine, University Medicine Greifswald, 17475 Greifswald
43. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Greifswald, Germany
44. Interfaculty Institute for Genetics and Functional Genomics; University Medicine and Ernst-Moritz-Arndt-University Greifswald, 17475 Greifswald, Germany
45. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, CB1 8RN
46. Centre for Cancer Genetic Epidemiology, Department of Oncology, Cambridge Centre, University of Cambridge, Cambride, UK, CB1 8RN
47. Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, 1683 Nicosia, Cyprus
48. Estonian Genome Center, University of Tartu, Tartu, Estonia
49. Department of Cardiovascular Sciences, University of Leicester, Cardiovascular Research Centre, Glenfield Hospital, Leicester, LE3 9QP, UK
50. NJS is supported by the British Heart Foundation and NJS is a NIHR Senior Investigator
51. School of Medicine, University of Dundee, Dundee, UK
52. Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK.
53. Division of Respiratory Medicine, University of Nottingham, Nottingham, UK
54. Population Health Research Institute, St George!s, University of London, London SW17 0RE, UK
55. University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands
56. Department of Medicine, Skellefteå Hospital, Skellefteå, Sweden
57. Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA
58. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ UK
59. Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA UK

60. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah 21589, Saudi Arabia

[±]Understanding Society: The UK Household Longitudinal Study: Michaela Benzeval[a], Jonathan Burton[a], Nicholas Buck[a], Annette Jäckle[a], Meena Kumari[a], Heather Laurie[a], Peter Lynn[a], Stephen Pudney[a], Birgitta Rabe[a], Dieter Wolke[b].

a. Institute for Social and Economic Research, University of Essex, UK
b. University of Warwick, UK

Consortium for Genetics of Smoking Behaviour Funding statements

Understanding Society Scientific Group is funded by the Economic and Social Research Council (ES/H029745/1) and the Wellcome Trust (WT098051). Paul D.P. Pharoah is funded by Cancer Research UK (C490/A16561). SHIP is funded by the German Federal Ministry of Education and Research (BMBF) and the German Research Foundation (DFG); see acknowledgements for details. F.W. Asselbergs is funded by the Netherlands Heart Foundation (2014T001) and UCL Hospitals NIHR Biomedical Research Centre. The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. Niek Verweij is supported by Horizon 2020 (Marie Sklodowska-Curie, 661395) and ICIN-NHI. LBC1921 and LBC1936 is supported by the MRC (MR/K026992/1). Paul W. Franks is supported by Novo Nordisk, the Swedish Research Council, Påhlssons Foundation, Swedish Heart Lung Foundation (2020389), and Skåne Regional Health Authority. Nicholas J Wareham, Claudia Langenberg, Robert A Scott, and Jian'an Luan are supported by the MRC (MC_U106179471 and MC_UU_12015/1). John C. Chambers and Jaspal S. Kooner are supported by the British Heart Foundation (SP/04/002), Medical Research Council (G0601966 and G0700931), Wellcome Trust (084723/Z/08/Z), NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143), Action on Hearing Loss (G51), National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, and iHEalth-T2D (643774). The BRIGHT study was supported by the Medical Research Council of Great Britain (Grant Number G9521010D); and by the British Heart Foundation (Grant Number PG/02/128). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. The Exome Chip genotyping was funded by Wellcome Trust Strategic Awards (083948 and 085475). We would also like to thank the Barts Genome Centre staff for their assistance with this project. The ASCOT study and the collection of the ASCOT DNA repository was supported by Pfizer, New York, NY, USA, Servier Research Group, Paris, France; and by Leo Laboratories, Copenhagen, Denmark. Genotyping of the Exome Chip in ASCOT-SC and ASCOT-UK was funded by the National Institutes of Health Research (NIHR). Anna F. Dominiczak was supported by the British Heart Foundation (Grant Numbers RG/07/005/23633, SP/08/005/25115); and by the European Union Ingenious HyperCare Consortium: Integrated Genomics, Clinical Research, and Care in Hypertension (grant number LSHM-C7-2006-037093). Nilesh J. Samani is supported by the British Heart Foundation. Panos Deloukas is supported by the British Heart Foundation (RG/14/5/30893), and NIHR, where his work forms part of the research themes contributing to the translational research portfolio of Barts Cardiovascular Biomedical Research Centre which is funded by the National Institute for Health Research (NIHR).

Consortium for Genetics of Smoking Behaviour Acknowledgements

The authors would like to thank the many colleagues who contributed to collection and phenotypic characterisation of the clinical samples, as well as genotyping and analysis of the GWA data. Special mentions are as follows:

Some of the data utilised in this study were provided by the Understanding Society: The UK Household Longitudinal Study, which is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The data were collected by NatCen and the genome wide scan data were analysed by the Wellcome Trust Sanger Institute. The Understanding Society DAC have an application system for genetics data and all use of the data should be approved by them. The application form is at: https://www.understandingsociety.ac.uk/about/health/data.

SHIP (Study of Health in Pomerania) and SHIP-TREND both represent population-based studies. SHIP is supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung (BMBF); grants 01ZZ9603, 01ZZ0103, and 01ZZ0403) and the German Research Foundation (Deutsche Forschungsgemeinschaft (DFG); grant GR 1912/5-1). SHIP and SHIP-TREND are part of the Community Medicine Research net (CMR) of the Ernst-Moritz-Arndt University Greifswald (EMAU) which is funded by the BMBF as well as the Ministry for Education, Science and Culture and the Ministry of Labor, Equal Opportunities, and Social Affairs of the Federal State of Mecklenburg-West Pomerania. The CMR encompasses several research projects that share data from SHIP. The EMAU is a member of the Center of Knowledge Interchange (CKI) program of the Siemens AG. SNP typing of SHIP and SHIP-TREND using the Illumina Infinium HumanExome BeadChip (version v1.0) was supported by the BMBF (grant 03Z1CN22).

LifeLines authors thank Behrooz Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Ilja Nolte, Lude Franke, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, Judith Vonk for their work related to data-collection and validation. The authors are grateful to the study participants, the staff from the LifeLines Cohort Study and Medical Biobank Northern Netherlands, and the participating general practitioners and pharmacists. LifeLines Scientific Protocol Preparation: Rudolf de Boer, Hans Hillege, Melanie van der Klauw, Gerjan Navis, Hans Ormel, Dirkje Postma, Judith Rosmalen, Joris Slaets, Ronald Stolk, Bruce Wolffenbuttel; LifeLines GWAS Working Group: Behrooz Alizadeh, Marike Boezen, Marcel Bruinenberg, Noortje Festen, Lude Franke, Pim van der Harst, Gerjan Navis, Dirkje Postma, Harold Snieder, Cisca Wijmenga, Bruce Wolffenbuttel. The authors wish to acknowledge the services of the LifeLines Cohort Study, the contributing research centres delivering data to LifeLines, and all the study participants.

Fenland authors thank Fenland Study volunteers for their time and help, Fenland Study general Practitioners and practice staff for assistance with recruitment, and Fenland Study Investigators, Co-ordination team and the Epidemiology Field, Data and Laboratory teams for study design, sample/data collection and genotyping.

We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study. In particular we thank Clare Muckian and David Toomey for their help in DNA extraction, storage, and handling. We would also like to acknowledge the Barts and The London Genome Centre staff for genotyping the Exome Chip array.

The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. We would also like to thank the Barts Genome Centre staff for their assistance with this project.

Patricia B. Munroe, Mark J. Caulfield, and Helen R. Warren wish to acknowledge the NIHR Cardiovascular Biomedical Research Unit at Barts and The London, Queen Mary University of

London, UK for support. Nilesh J. Samani and Mark J. Caulfield are Senior National Institute for Health Research Investigators.
EMBRACE Collaborating Centres are: Coordinating Centre, Cambridge: Daniel Barrowdale, Debra Frost, Jo Perkins. North of Scotland Regional Genetics Service, Aberdeen: Zosia Miedzybrodzka, Helen Gregory. Northern Ireland Regional Genetics Service, Belfast: Patrick Morrison, Lisa Jeffers. West Midlands Regional Clinical Genetics Service, Birmingham: Kai-ren Ong, Jonathan Hoffman. South West Regional Genetics Service, Bristol: Alan Donaldson, Margaret James. East Anglian Regional Genetics Service, Cambridge: Joan Paterson, Marc Tischkowitz, Sarah Downing, Amy Taylor. Medical Genetics Services for Wales, Cardiff: Alexandra Murray, Mark T. Rogers, Emma McCann. St James's Hospital, Dublin & National Centre for Medical Genetics, Dublin: M. John Kennedy, David Barton. South East of Scotland Regional Genetics Service, Edinburgh: Mary Porteous, Sarah Drummond. Peninsula Clinical Genetics Service, Exeter: Carole Brewer, Emma Kivuva, Anne Searle, Selina Goodman, Kathryn Hill. West of Scotland Regional Genetics Service, Glasgow: Rosemarie Davidson, Victoria Murday, Nicola Bradshaw, Lesley Snadden, Mark Longmuir, Catherine Watt, Sarah Gibson, Eshika Haque, Ed Tobias, Alexis Duncan. South East Thames Regional Genetics Service, Guy's Hospital London: Louise Izatt, Chris Jacobs, Caroline Langman. North West Thames Regional Genetics Service, Harrow: Huw Dorkins. Leicestershire Clinical Genetics Service, Leicester: Julian Barwell. Yorkshire Regional Genetics Service, Leeds: Julian Adlard, Gemma Serra-Feliu. Cheshire & Merseyside Clinical Genetics Service, Liverpool: Ian Ellis, Claire Foo. Manchester Regional Genetics Service, Manchester: D Gareth Evans, Fiona Lalloo, Jane Taylor. North East Thames Regional Genetics Service, NE Thames, London: Lucy Side, Alison Male, Cheryl Berlin. Nottingham Centre for Medical Genetics, Nottingham: Jacqueline Eason, Rebecca Collier. Northern Clinical Genetics Service, Newcastle: Alex Henderson, Oonagh Claber, Irene Jobson. Oxford Regional Genetics Service, Oxford: Lisa Walker, Diane McLeod, Dorothy Halliday, Sarah Durell, Barbara Stayner. The Institute of Cancer Research and Royal Marsden NHS Foundation Trust: Ros Eeles, Nazneen Rahman, Elizabeth Bancroft, Elizabeth Page, Audrey Ardern-Jones, Kelly Kohut, Jennifer Wiggins, Jenny Pope, Sibel Saya, Natalie Taylor, Zoe Kemp and Angela George. North Trent Clinical Genetics Service, Sheffield: Jackie Cook, Oliver Quarrell, Cathryn Bardsley. South West Thames Regional Genetics Service, London: Shirley Hodgson, Sheila Goff, Glen Brice, Lizzie Winchester, Charlotte Eddy, Vishakha Tripathi, Virginia Attard. Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton: Diana Eccles, Anneke Lucassen, Gillian Crawford, Donna McBride, Sarah Smalley.

Consortium for Genetics of Smoking Behaviour Conflict of Interest statements

Paul W. Franks has been a paid consultant for Eli Lilly and Sanofi Aventis and has received research support from several pharmaceutical companies as part of European Union Innovative Medicines Initiative (IMI) projects. Neil Poulter has received financial support from several pharmaceutical companies that manufacture either blood pressure lowering or lipid lowering agents or both and consultancy fees. Peter Sever has received research awards from Pfizer. Mark J. Caulfield is Chief Scientist for Genomics England, a UK government company.

References

1. Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, Coll CRA (2002): Selected major risk factors and global and regional burden of disease. *Lancet*. 360:1347-1360.
2. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. (2015): Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*.
3. Eng MY, Luczak SE, Wall TL (2007): ALDH2, ADH1B, and ADH1C genotypes in Asians: A literature review. *Alcohol Res Health*. 30:22-27.
4. Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, et al. (2010): Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genet*. 42:441-U134.
5. Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, et al. (2010): Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet*. 6.
6. Bierut LJ, Stitzel JA (2014): Genetic Contributions of the alpha 5 Nicotinic Receptor Subunit to Smoking Behavior. *Recept Ser*. 26:327-339.
7. Luczak SE, Glatt SJ, Wall TL (2006): Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians. *Psychol Bull*. 132:607-621.
8. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. (2016): Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536:285-+.
9. Minikel E, Lek M, Samocha KE, Karczewski KJ, Marshall JL, Armean I, et al. (2016): An early glimpse of saturation mutagenesis in humans: Insights from protein-coding genetic variation in 60,706 people. *Prion*. 10:S107-S107.
10. Sveinbjornsson G, Albrechtsen A, Zink F, Gudjonsson SA, Oddson A, Masson G, et al. (2016): Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet*. 48:314-317.
11. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, et al. (2017): Rare and low-frequency coding variants alter human adult height. *Nature*. 542:186-190.
12. Yang J, Wang S, Yang Z, Hodgkinson CA, Iarikova P, Ma JZ, et al. (2014): The contribution of rare and common variants in 30 genes to risk nicotine dependence. *Mol Psychiatry*.
13. McClure-Begley TD, Papke RL, Stone KL, Stokes C, Levy AD, Gelernter J, et al. (2014): Rare human nicotinic acetylcholine receptor alpha4 subunit (CHRNA4) variants affect expression and function of high-affinity nicotinic acetylcholine receptors. *The Journal of pharmacology and experimental therapeutics*. 348:410-420.
14. Piliguian M, Zhu AZ, Zhou Q, Benowitz NL, Ahluwalia JS, Sanderson Cox L, et al. (2014): Novel CYP2A6 variants identified in African Americans are associated with slow nicotine metabolism in vitro and in vivo. *Pharmacogenet Genomics*. 24:118-128.
15. Haller G, Druley T, Vallania FL, Mitra RD, Li P, Akk G, et al. (2012): Rare missense variants in CHRNB4 are associated with reduced risk of nicotine dependence. *Hum Mol Genet*. 21:647-655.
16. Haller G, Kapoor M, Budde J, Xuei X, Edenberg H, Nurnberger J, et al. (2014): Rare missense variants in CHRNB3 and CHRNA3 are associated with risk of alcohol and cocaine dependence. *Hum Mol Genet*. 23:810-819.
17. Zuo L, Tan Y, Li C-SR, Wang Z, Wang K, Zhang X, et al. (2016): Associations of rare nicotinic cholinergic receptor gene variants to nicotine and alcohol dependence. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*.
18. Xie P, Kranzler HR, Krauthammer M, Cosgrove KP, Oslin D, Anton RF, et al. (2011): Rare Nonsynonymous Variants in Alpha-4 Nicotinic Acetylcholine Receptor Gene Protect Against Nicotine Dependence. *Biological Psychiatry*. 70:528-536.

19. Wessel J, McDonald SM, Hinds Da, Stokowski RP, Javitz HS, Kennemer M, et al. (2010): Resequencing of Nicotinic Acetylcholine Receptor Genes and Association of Common and Rare Variants with the Fagerström Test for Nicotine Dependence. *Neuropsychopharmacology*. 35:2392-2402.
20. Thorgeirsson TE, Steinberg S, Reginsson GW, Bjornsdottir G, Rafnar T, Jonsdottir I, et al. (2016): A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. *Molecular Psychiatry*. 21:594-600.
21. Olfson E, Saccone NL, Johnson EO, Chen L-S, Culverhouse R, Doheny K, et al. (2016): Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans. *Molecular Psychiatry*. 21:601-607.
22. Peng Q, Gizer IR, Libiger O, Bizon C, Wilhelmsen KC, Schork NJ, et al. (2014): Association and ancestry analysis of sequence variants in ADH and ALDH using alcohol-related phenotypes in a Native American community sample. *Am J Med Genet B Neuropsychiatr Genet*. 165B:673-683.
23. Way M, McQuillin A, Saini J, Ruparelia K, Lydall GJ, Guerrini I, et al. (2015): Genetic variants in or near ADH1B and ADH1C affect susceptibility to alcohol dependence in a British and Irish population. *Addiction Biology*. 20:594-604.
24. Zuo L, Wang K-S, Zhang X-Y, Li C-sR, Zhang F, Wang X, et al. (2013): Rare SERINC2 variants are specific for alcohol dependence in individuals of European descent. *Pharmacogenetics and Genomics*. 23:395-402.
25. Zuo L, Zhang X, Deng H-w, Luo X (2013): Association of rare PTP4A1-PHF3-EYS variants with alcohol dependence. *Journal of Human Genetics*. 58:178-179.
26. Haller G, Li P, Esch C, Hsu S, Goate AM, Steinbach JH (2014): Functional Characterization Improves Associations between Rare Non-Synonymous Variants in CHRNB4 and Smoking Behavior. *PLoS ONE*. 9:e96753.
27. Duncan LE, Keller MC (2011): A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. *Am J Psychiat*. 168:1041-1049.
28. Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis GR, et al. (2014): Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biol Psychiatry*. 75:783-789.
29. Vrieze SI, Malone SM, Vaidyanathan U, Kwong A, Kang HM, Zhan X, et al. (2014): In search of rare variants: preliminary results from whole genome sequencing of 1,325 individuals with psychophysiological endophenotypes. *Psychophysiology*. 51:1309-1320.
30. Vrieze SI, Malone SM, Pankratz N, Vaidyanathan U, Miller MB, Kang HM, et al. (2014): Genetic associations of nonsynonymous exonic variants with psychophysiological endophenotypes. *Psychophysiology*. 51:1300-1308.
31. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. (2016): A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*. 48:1279-1283.
32. Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, et al. (2018): Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nature Genet*. 50:737-+.
33. Schaid DJ, Chen W, Larson NB (2018): From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 19:491-504.
34. Hicks BM, Schalet BD, Malone SM, Iacono WG, McGue M (2011): Psychometric and genetic architecture of substance use disorder and behavioral disinhibition measures for gene association studies. *Behav Genet*. 41:459-475.
35. Vrieze SI, McGue M, Miller MB, Hicks BM, Iacono WG (2013): Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence: twin biometry, GCTA, and genome-wide scoring. *Behav Genet*. 43:97-107.

36. Vink JM, Willemsen G, Boomsma DI (2005): Heritability of smoking initiation and nicotine dependence. *Behav Genet*. 35:397-406.
37. Maes HH, Sullivan PF, Bulik CM, Neale MC, Prescott CA, Eaves LJ, et al. (2004): A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychological Medicine*. 34:1251-1261.
38. Swan GE, Carmelli D, Rosenman RH, Fabsitz RR, Christian JC (1990): Smoking and alcohol consumption in adult male twins: genetic heritability and shared environmental influences. *J Subst Abuse*. 2:39-50.
39. Schumann G, Liu CY, O'Reilly P, Gao H, Song P, Xu B, et al. (2016): KLB is associated with alcohol drinking, and its gene product beta-Klotho is necessary for FGF21 regulation of alcohol preference. *P Natl Acad Sci USA*. 113:14372-14377.
40. Jorgenson E, Thai KK, Hoffmann TJ, Sakoda LC, Kvale MN, Banda Y, et al. (2017): Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Mol Psychiatry*.
41. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. (2010): Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genet*. 42:448-U135.
42. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. (2017): Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*.
43. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. (2010): Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 42:348-354.
44. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, et al. (2014): Meta-analysis of gene-level tests for rare variant association. *Nat Genet*. 46:200-204.
45. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. (2014): RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res*. 42:D756-763.
46. Wu MC, Lee S, Cai TX, Li Y, Boehnke M, Lin XH (2011): Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *Am J Hum Genet*. 89:82-93.
47. Lee S, Wu MC, Lin X (2012): Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*. 13:762-775.
48. Li B, Leal SM (2008): Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*. 83:311-321.
49. Jiang Y, Chen S, McGuire D, Chen F, Liu M, Iacono WG, et al. (2018): Proper conditional analysis in the presence of missing data: Application to large scale meta-analysis of tobacco use phenotypes. *PLoS genetics*. 14:e1007452.
50. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. (2018): Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *bioRxiv*.
51. Pickrell JK (2014): Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet*. 94:559-573.
52. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*. 47:291-295.
53. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genet*. 47:291-+.
54. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. (2015): An atlas of genetic correlations across human diseases and traits. *Nature Genet*. 47:1236-+.

55. Zhan X, Liu DJ (2015): SEQMINER: An R-Package to Facilitate the Functional Interpretation of Sequence-Based Associations. *Genet Epidemiol*.
56. Lassi G, Taylor AE, Timpson NJ, Kenny PJ, Mather RJ, Eisen T, et al. (2016): The CHRNA5-A3-B4 Gene Cluster and Smoking: From Discovery to Therapeutics. *Trends Neurosci*. 39:851-861.
57. Edenberg HJ (2007): The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health*. 30:5-13.
58. Ghitza UE, Zhai H, Wu P, Airavaara M, Shaham Y, Lu L (2010): Role of BDNF and GDNF in drug reward and relapse: a review. *Neurosci Biobehav Rev*. 35:157-171.
59. Paul S, Amundson SA (2014): Differential Effect of Active Smoking on Gene Expression in Male and Female Smokers. *J Carcinog Mutagen*. 5.

Table 1. All nonsynonymous/loss of function variants with posterior probability of association > .80 from one of the two fine mapping methods.

SNP	REF/ALT	N	ALT AF	GWAS p-value	Beta	SE	Direction	Annotation	Posterior Probability of Association		Number SNPs (Low Frequency Coding SNPs) in 95% Credible Interval	
									W/out Functional Prior	W/ Functional Prior (fgwas)	W/out Functional Prior	W/ Functional Prior (fgwas)
Cigarettes per Day (CigDay)												
rs36015615[a]	G/A	69,951	.0002	3.2×10^{-8}	1.2	.210	==+=++=+=X=X+X+++	Nonsynonymous [*STARD3*]	.82	.62	8,997 (6211)	11302 (6232)
rs16969968	G/A	153,918	.34	2.5×10^{-139}	.096	.0038	+-++-+++++-++++++	Nonsynonymous [*CHRNA5*]	.84	.92	2(0)	2 (0)
Drinks per Week (DrnkWk)												
rs1260326	T/C	357,854	.61	4.6×10^{-40}	0.032	.0024	+++++++++--++++++	Nonsynonymous [*GCKR*]	1.0	1.0	1 (0)	1 (0)
rs1229984	T/C	334,588	.98	2.3×10^{-173}	0.25	.0088	=+-XXXX+XXXX=++++	Nonsynonymous [*ADH1B*]	1.0	1.0	1 (1)	1 (1)
rs28929474	C/T	357,854	.02	2.2×10^{-11}	-0.057	.0085	--++-----+++----+	Nonsynonymous [*SERPINA1*]	.95	1.0	1 (1)	1 (1)
rs1800566	G/A	357,854	.18	2.00×10^{-8}	0.017	.0031	++++++++++--+++-	Nonsynonymous [*NQO1*]	.32	.97	103 (0)	1 (0)
Smoking Initiation (SmkInit)												
rs2232423	A/G	433,216	.11	1.40×10^{-8}	-0.019	.0034	-+-+---+------	Nonsynonymous [*ZSCAN12*]	.84	.64	502 (0)	2 (0)
rs35891966	G/A	433,216	.07	1.30×10^{-8}	-0.024	.0042	---+-+---++---	Nonsynonymous [*NAV2*]	.98	1.0	1 (0)	1 (0)
rs147052174	G/T	433,216	.02	1.2×10^{-7}	.043	.0080	+++++++++-++++	Nonsynonymous [*FAM163A*]	.81	1.0	2432(66)	1 (0)
rs6265	C/T	433,216	.19	1.9×10^{-10}	-.017	.0030	++-++-++--+---	Nonsynonymous [*BDNF*]	.32	.83	25(0)	2 (0)
rs61754158	C/T	433,216	.01	1.4×10^{-6}	-.055	.0114	--+-+---=+=-+-	Nonsynonymous [*HEATR5A*]	.39	.87	9742(195)	9742 (195)
rs34967813	A/G	433,216	.31	8.1×10^{-7}	-.011	.0023	----+--+--+---	Nonsynonymous [*RYR2*]	.14	.98	7413(56)	1 (0)

[a]rs36015615 did not replicate in two additional datasets. See results section.

Note: REF=reference allele on GRCh37, ALT=alternate allele, N=sample size across all studies that genotyped the variant, ALT AF=allele frequency of the alternate allele estimated in the meta-analysis. A variant is considered “rare” if MAF < .01, and low frequency if .01 ≤ MAF < .05. In the DIRECTION column each symbol represents the contribution of one of the studies. A “+” indicates the ALT allele had a positive effect in that study; “-“ indicates a negative effect. A “=” indicates the variant was monomorphic and “X” indicates it was absent in that study. The order of studies for CigDay and DrnkWk was ARIC, UKB, COGA, FINNTWIN, FUSION, GECCO, HRS, ID1000, MEC, METSIM, MHI, MCTFR, NAGOZALC, NESCOG, SardiNIA, TwinsUK, and WHI. For SmkInit the order is the same except COGA and MCTFR were not available. See the supplemental materials for study acronym explanations.

Table 2. Significant gene based test results, assuming a Bonferroni threshold of .05/20,000=2.5×10^{-6}.

Phenotype	Gene	N	Number Variants	Beta	SE	p-value	Method
Age of Initiation of Smoking	*ARHGEF37*	147,010	17	.08	.017	1.9×10^{-6}	Burden
Smoking Initiation	*HEATR5A*	427,262	41	-.02	.009	1.4×10^{-8}	SKAT
Drinks per Week	*ADH1C*	353,265	4	-.15	.014	1.8e-27	Burden
Drinks per Week	*ADH1C*	353,265	4	-.15	.014	1.9e-40	SKAT

Note: no significant genes were identified for the other two phenotypes.

Table 3: Estimation of Heritability Explained by Variants on Exome Array. We estimate the heritability based upon a baseline model with seven different functional categories. The reported heritability $\hat{h}^2$ is based upon the cumulative value from the functional categories with significant heritabilities. We also report the standard deviation ($se(\hat{h}^2)$) and p-values, estimated using jackknife.

Annotation	**Phenotype**	**Heritability Estimates**		
		$\widehat{\boldsymbol{h}}^{\boldsymbol{2}}$	$\boldsymbol{se}(\widehat{\boldsymbol{h}}^{\boldsymbol{2}})$	**P-Value**
All Variants	Age of Initiation of smoking	.06	.0049	7.7×10^{-35}
	Cigarettes per Day	.09	.0019	$< 2.2\times10^{-303}$
	Pack Years	.10	.0022	$< 2.2\times10^{-303}$
	Smoking Initiation	.14	.0007	$< 2.2\times10^{-303}$
	Drinks per Week	.16	.0089	7.3×10^{-73}
Rare Coding Variants (MAF<.01)	Age of Initiation of smoking	.011	.0015	2.8×10^{-2}
	Cigarettes per Day	.010	.0006	1.7×10^{-2}
	Pack Years	.018	.0007	8.5×10^{-6}
	Smoking Initiation	.022	.0002	3.9×10^{-16}
	Drinks per Week	.020	.0013	1.8×10^{-7}

Table 4: Estimation of Genetic Correlation Between Smoking and Drinking Traits. We estimate genetic correlations between five smoking and drinking traits. Genetic correlation estimates ($\hat{r}_g$), their standard deviation ($se(\hat{r}_g)$) and p-values are reported.

Trait 1	Trait 2	Genetic Correlation		
		$\hat{r}_g$	$se(\hat{r}_g)$	P-value
A. Aggregated Genetic Correlation Induced by All Variants on the Exome Array				
Drinks per Week	Smoking Initiation	.43	.06	1.7×10^{-11}
Drinks per Week	Age of Initiation of Smoking	.01	.13	9.3×10^{-1}
Drinks per Week	Pack Years	.22	.10	2.6×10^{-2}
Drinks per Week	Cigarettes per Day	.20	.09	3.1×10^{-2}
Smoking Initiation	Age of Initiation of Smoking	-.64	.11	1.1×10^{-8}
Smoking Initiation	Pack Years	.45	.08	4.9×10^{-8}
Smoking Initiation	Cigarettes per Day	.10	.07	1.5×10^{-1}
Age of Initiation of Smoking	Pack Years	-.63	.17	2.1×10^{-4}
Age of Initiation of Smoking	Cigarettes per Day	-.26	.16	9.9×10^{-2}
Pack Years	Cigarettes per Day	.77	.13	2.2×10^{-9}
B. Genetic Correlation Induced by Rare (MAF < 1%) Nonsynonymous Variants				
Drinks per Week	Smoking Initiation	.49	.08	1.2×10^{-10}
Drinks per Week	Age of Initiation of Smoking	-.04	.30	8.9×10^{-1}
Drinks per Week	Pack Years	.08	.02	2.7×10^{-4}
Drinks per Week	Cigarettes per Day	.09	.02	5.2×10^{-5}
Smoking Initiation	Age of Initiation of Smoking	-1.10	.21	1.3×10^{-7}
Smoking Initiation	Pack Years	.63	.08	1.5×10^{-14}
Smoking Initiation	Cigarettes per Day	.23	.08	3.3×10^{-3}
Age of Initiation of Smoking	Pack Years	-1.10	.33	1.5×10^{-3}
Age of Initiation of Smoking	Cigarettes per Day	-.69	.32	3.2×10^{-2}
Pack Years	Cigarettes per Day	.87	.14	1.4×10^{-9}

Exome Chip Meta-Analysis Fine Maps Causal Variants and Elucidates the Genetic Architecture of Rare Coding Variants in Smoking and Alcohol Use

Supplement 1

Complete summary statistics

Complete sets of summary statistics are available for download here: https://genome.psych.umn.edu/index.php/GSCAN.

Analysis plan

The analysis plan used by all studies to generate summary statistis is here: https://genome.psych.umn.edu/index.php/GSCAN.

Exploratory analyses of individuals of African ancestry

Using the same techniques as for individuals of European ancestry, we conducted a GWAS meta-analysis of three cohorts of African and African admixed ancestry. These cohorts were the UK Biobank, The Collaborative Study on the Genetics of Alcoholism (COGA), and the Health and Retirement Study (HRS). Sample sizes and genomic controls are provided in **Table S3**. African ancestry in the UK Biobank were identified through inspection of genetic principal component 1 against component 2. Individuals with values PC2>0 and PC1>150.

Ultimately, one genome-wide significant hit (rs3806243, p=2.3×10^{-8}) was associated with cigarettes per day at the conventional African-ancestry $p < 2.5\times10^{-8}$ threshold. This locus had not been discovered in a prior larger meta-analysis of cigarettes per day in African American individuals [1]. Given the lack of replication in the larger sample and marginal statistical evidence, no further analyses were conducted. We encourage investigators to continue to build cohorts of non-European ancestry. QQ plots and Manhattan plots are provided in **Figures S3 and S4.**

Sources of funding of individual studies

COGA: The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, H. Edenberg, L. Bierut, includes eleven different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, J. Nurnberger Jr., T. Foroud); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St. Louis (L. Bierut, J. Rice, K. Bucholz, A. Agrawal); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield, A. Brooks); Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia; Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA (L. Almasy), Virginia Commonwealth University (D. Dick), Icahn School of Medicine at Mount Sinai (A. Goate), and Howard University (R. Taylor). Other COGA collaborators include: L. Bauer (University of Connecticut); J. McClintick, L. Wetherill, X. Xuei, Y. Liu, D. Lai, S. O'Connor, M. Plawecki, S. Lourens (Indiana University); G. Chan (University of Iowa; University of Connecticut); J. Meyers, D. Chorlian, C. Kamarajan, A. Pandey, J. Zhang (SUNY Downstate); J.-C. Wang, M. Kapoor, S. Bertelsen (Icahn School of Medicine at Mount Sinai); A. Anokhin, V. McCutcheon, S. Saccone (Washington University); J. Salvatore, F. Aliev, B. Cho (Virginia Commonwealth University); and Mark Kos (University of Texas Rio Grande Valley). A. Parsian and M. Reilly are the NIAAA Staff Collaborators.

We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA).

FTC: Phenotyping and genotyping of the Finnish Twin Cohort (FTC) has been supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (grants 213506, 129680), the Academy of Finland (grants 100499, 205585, 118555, 141054, 265240, 263278 and 264146 to J. Kaprio), National Institute for Health (grant DA12854 to P.A.F. Madden), National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to R. J. Rose and AA15416 and K02AA018755 to D. M. Dick), Sigrid Juselius Foundation (to J. Kaprio), Global Research Award for Nicotine Dependence, Pfizer Inc. (to J. Kaprio), and the Welcome Trust Sanger Institute, UK. Antti-Pekka Sarin and Samuli Ripatti are acknowledged for genotype data quality controls and imputation. Association analyses were run at the ELIXIR Finland node hosted at CSC – IT Center for Science for ICT resources.

GECCO: Support for this study came from the National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088; R01CA059045). The authors also thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible.

Substudies of GECCO:

ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in

this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

CPS-II: The authors thank the CPS-II participants and Study Management Group for their invaluable contributions to this research. The authors would also like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, and cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program.

HPFS, NHS: We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS under the supervision of Dr. Immaculata Devivo and Dr. David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS. We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Additionally, a subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS (Yeager, M et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007 May;39(5):645-9), CGEMS pancreatic cancer scan (PanScan) (Amundadottir, L et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009 Sep;41(9):986-90, and Petersen, GM et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010 Mar;42(3):224-8), and the Lung Cancer and Smoking study (Landi MT, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009 Nov;85(5):679-91). The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207.v1.p1 and phs000206.v3.p2, respectively, and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093.v2.p2. Funding for the Lung Cancer and Smoking study was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping. The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O'Brien and staff, Westat, Inc., and Drs. Bill Kopp and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

PMH: National Institutes of Health (R01 CA076366 to P.A. Newcomb). The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study.

CCFR: This work was supported by grant UM1 CA167551 from the National Cancer Institute and through cooperative agreements with the following CCFR centers: **Ontario Familial Colorectal Cancer Registry** (U01/U24 CA074783)

HRS: HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the University of Michigan School of Public Health.

MEC: Support for this study came from the National Institutes of Health (R37CA54281, P01CA033619, R01CA63464).

MCTFR: Data collection and analysis was supported by National Institutes of Health awards DA036216, DA05147, and DA024417.

MHI: We thank all participants and staff of the André and France Desmarais Montreal Heart Institute's (MHI) Biobank. The genotyping of the MHI Biobank was done at the MHI Pharmacogenomic Centre and funded by the MHI Foundation. Valerie Turcot is supported by a postdoctoral fellowship from the Canadian Institutes of Health Research (CIHR). Jean-Claude Tardif and Guillaume Lettre are supported by the Canada Research Chair Program.

NESCOG: This work is supported by the Netherlands Organization for Scientific Research (NWO Brain & Cognition 433-09-228, NWO Complexity Project 645-000-003, NWO VICI 453-14-005). Statistical analyses were carried out on the Genetic Cluster Computer hosted by SURFsara and financially supported by the Netherlands Organization for Scientific Research (NWO 480-05-003 PI: Posthuma) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam.

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. Personal funding for Sean P. David from National Institute on Minority Health and Health Disparities grant U54-MD010724. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.

Figure S1. QQ plots of GWAS meta-analysis in individuals of European ancestry

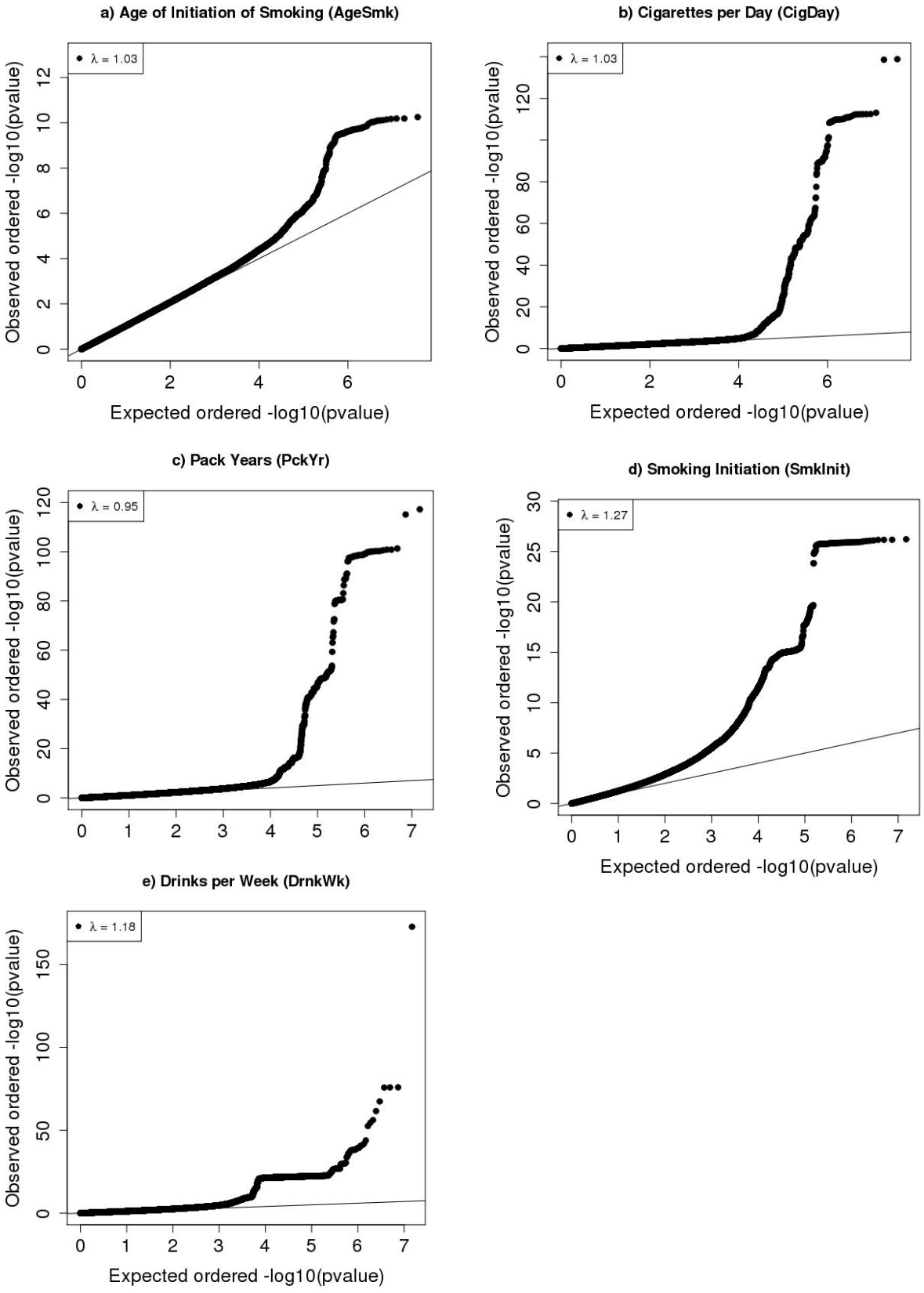
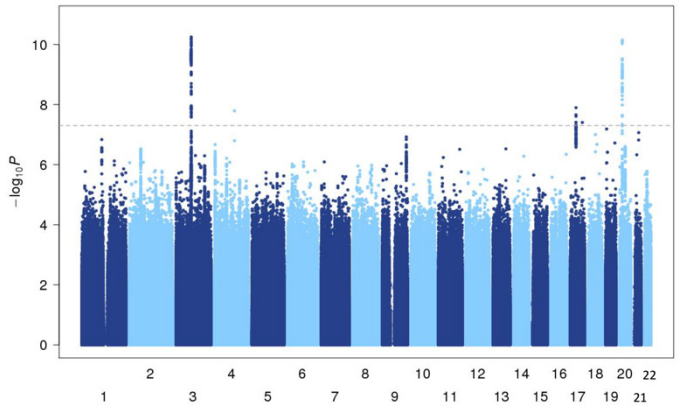
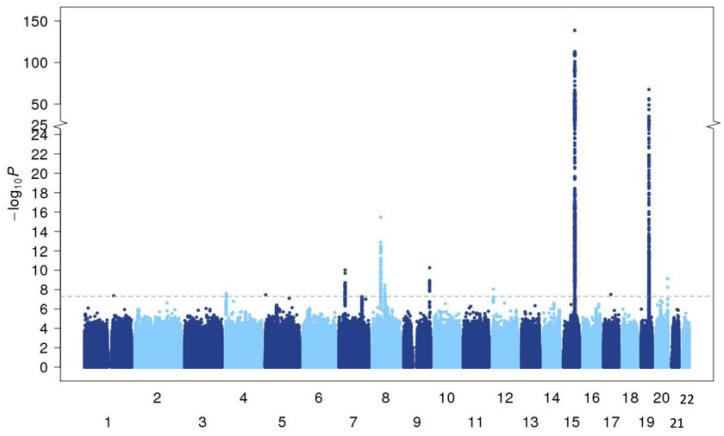


Figure S2. Manhattan plots of GWAS meta-analysis in individuals of European ancestry

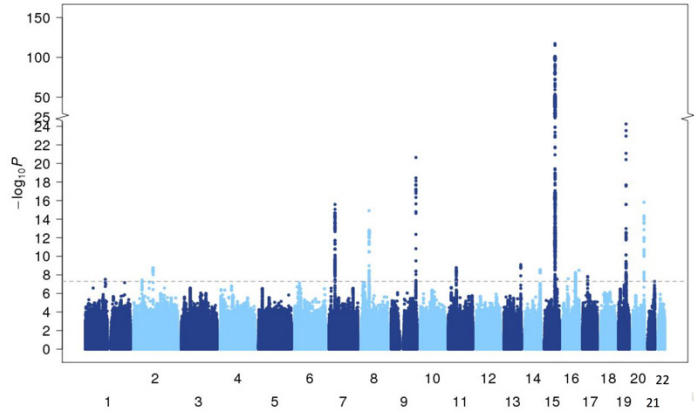
a) Age of Initiation of Regular Smoking (AgeSmk)



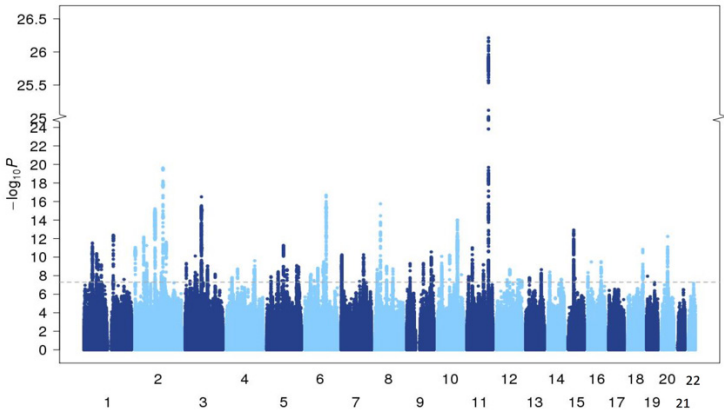
b) Cigarettes per Day (CigDay)



c) Pack Years (PckYr)



d) Smoking Initiation (SmkInit)



e) Drinks per Week (DrnkWk)

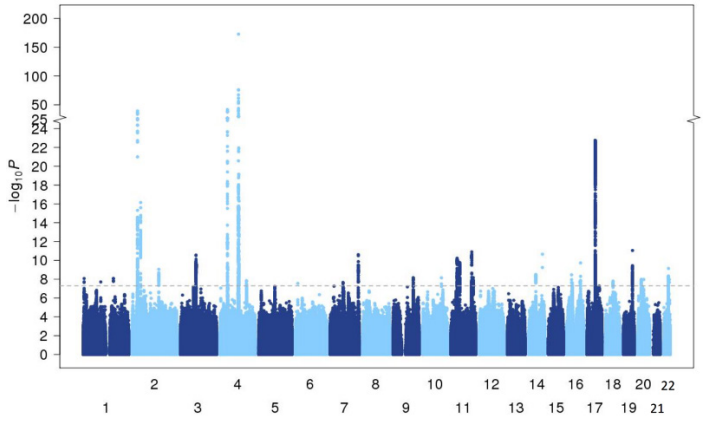


Figure S3. QQ plots of GWAS meta-analysis in individuals of African ancestry

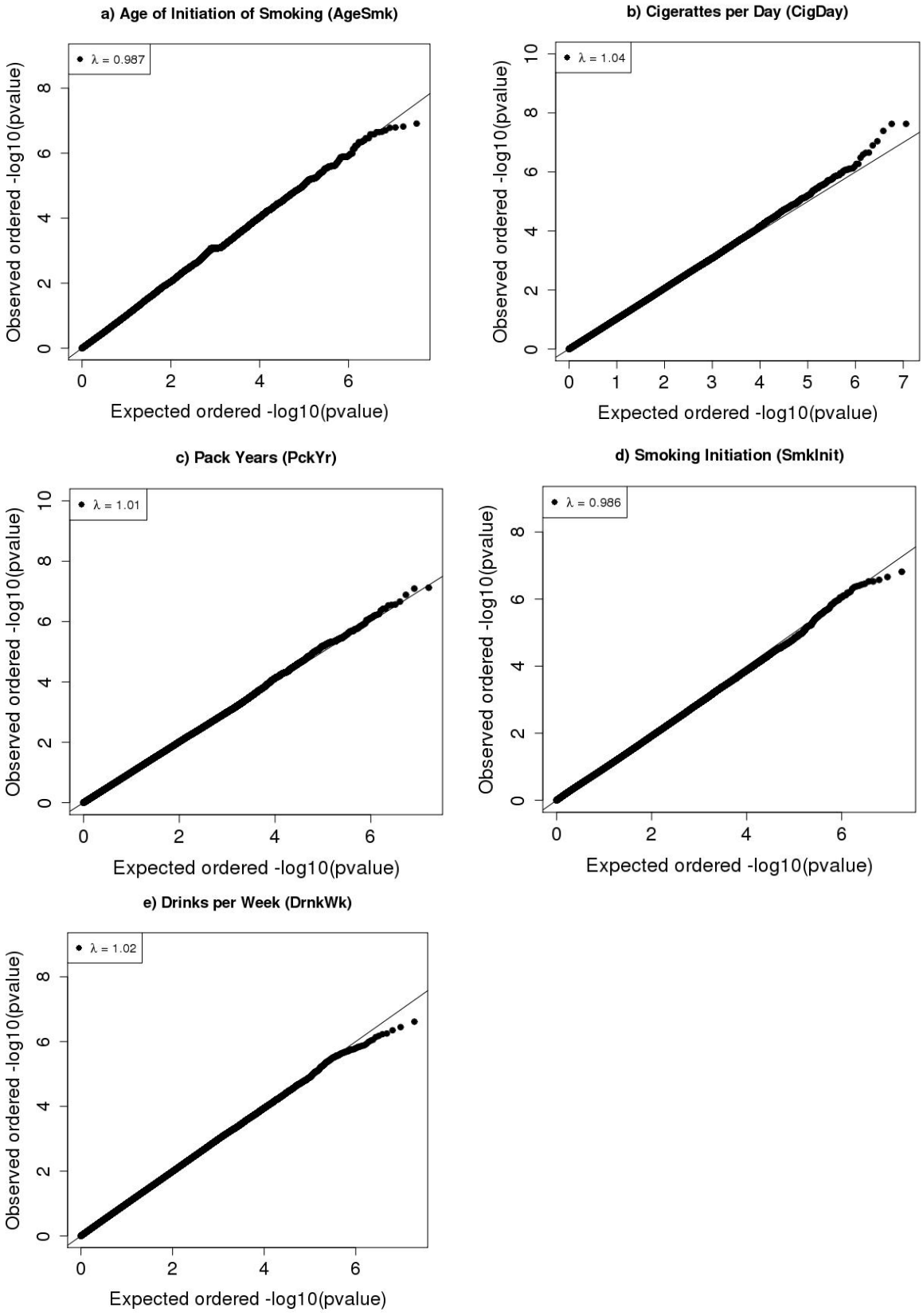
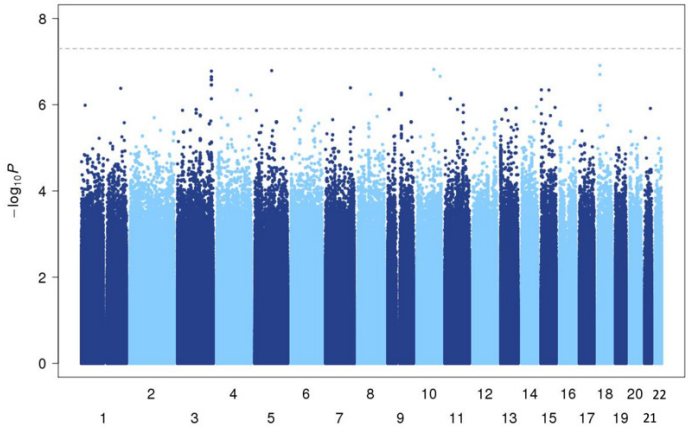
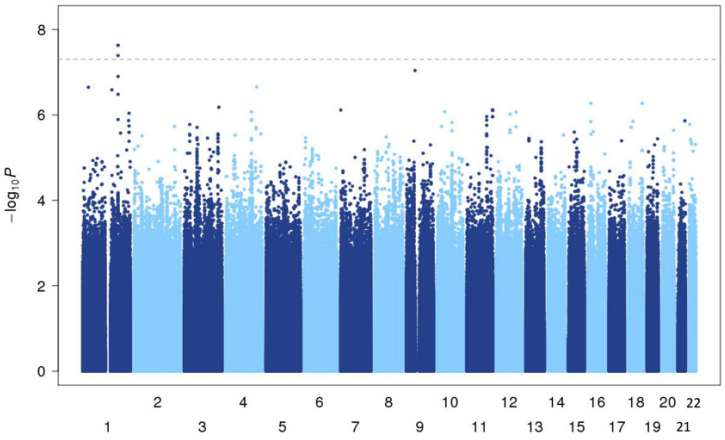


Figure S4. Manhattan plots of GWAS meta-analysis in individuals of African ancestry

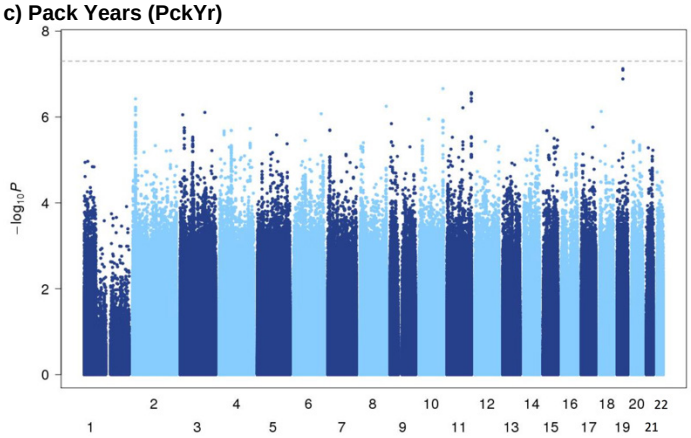
a) Age of Initiation of Regular Smoking (AgeSmk)



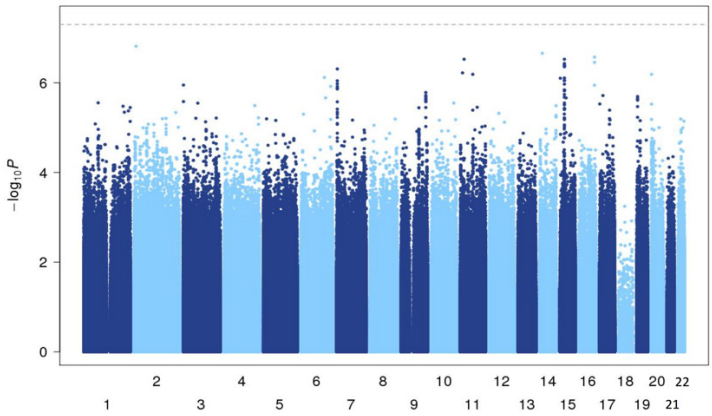
b) Cigarettes per Day (CigDay)



c) Pack Years (PckYr)



d) Smoking Initiation (SmkInit)



e) Drinks per Week (DrnkWk)

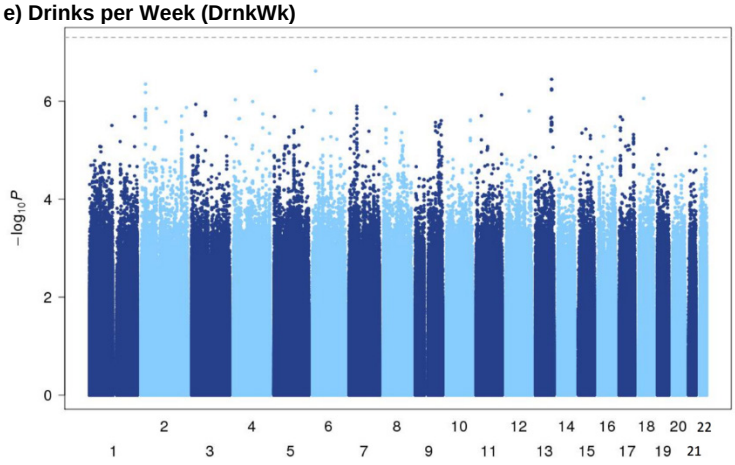


Table S1. Participating cohort descriptions

Study Abbreviation	Full Study Name	Design	Array Platform	Association Covariates
ARIC	Atherosclerosis Risk in Communities	Community sample of older adults	Illumina HumanExome	
COGA*	Collaborative Study on the Genetics of Alcoholism	Family study of alcoholism	Illumina HumanCoreExome*	Sex, age, sex*age, age^2, birth cohort, DSM5 alcohol dependence
FTC	NAG-FIN, FinnTwin12, FinnTwin16, FITSA	Population-based twin samples from the Older and Younger Finnish Twin Cohorts	Illumina HumanCoreExome	Sex, age, age^2, current or former smoker (cigarettes per day), year of birth, cohort status, BMI.
FUSION	Finland-United States Investigation of NIDDM Genetics	Type-2 diabetes case-control	Illumina HumanExome	Sex, age, age^2, current v. former smoker (cigarettes per day), height and weight (drinks per week).
GECCO	Genetics and Epidemiology of Colorectal Cancer Consortium	Colorectal cancer case-control	Illumina HumanExome	
HRS	Health and Retirement Study	National representative sample of older adults	Illumina HumanExome	Age, age^2, sex, age*sex, birth year, PCs 1-4 (European ancestry) or PCs 1-10 (African ancestry), current v. former smoker (for smoking outcomes), weight, bmi, bmi*gender, and current v. former drinker (for drinking outcomes)
ID1000	-	National representative sample of young adults	Illumina HumanExome	Age, age^2, sex, age*sex, PCs 1-10, current v. former smoker (for cigarettes/day, pack years); bmi, weight, height, bmi*sex for (drinks per week)
MEC	Multi-Ethnic Cohort		Illumina HumanExome	

Study Abbreviation	Full Study Name	Design	Array Platform	Association Covariates
METSIM	Metabolic Syndrome in Men		Illumina HumanExome	Sex, age, age^2, current v. former smoker (cigarettes per day), height and weight (drinks per week).
MHI	Montreal Heart Institute	Community sample of adults among visitors, patients and employees of the MHI.	Illumina HumanExome	Sex, age, age^2, PCs 1-10, current or former smoker status (for cigarettes per day), height and weight (for drinks per week).
MCTFR	Minnesota Center for Twin and Family Research	Community-based family cohort	Illumina HumanExome	Sex, age, parent-child generation,
NAGOZALC			Illumina HumanExome	
NESCOG	Netherlands study of Cognition, Environment, and Genes	National representative sample of adults	Illumina HumanExome	Age, age^2, sex, age*sex, PCs 1-10; current v. former smoker (for cigarettes/day, pack years); bmi, weight, height, bmi*sex (for drinks/week).
SardiNIA	-	Community-based Family Cohort	Illumina HumanExome	Sex, age, age^2, current v. former smoker (cigarettes per day), height and weight (drinks per week).
TwinsUK	-	Twin cohort	Illumina HumanExome	
WHI	Womens Health Initiative	Complex design consisting of clinical trials and observational cohort.	Illumina HumanExome	Sex, age, age^2, EV1, EV2, EV3 (all phenotypes); current v. former smoking (cigarettes per day), height and weight (drinks per week).
UK Biobank**	(Stratified by UK BiLEVE sample [N~50,000] and remainder).	Community sample of older adults, selected for heavy and non-smokers	UK BiLEVE and UK Axiom arrays	Sex, age, age^2, current or former smoker (for cigarettes per day), PCs 1-15, height, and weight (for drinking

*The exome array genotyping in COGA was performed in three broader groups comprised of 1059 founder subjects from 118 extended European American families and 2815 longitudinally ascertained subjects of mixed ethnicities. The 1059 subjects in 118 families were selected using the ExomePick program (http://genome.sph.umich.edu/wiki/ExomePicks) that uses the kinship information to suggests individuals to be sequenced in a large pedigree. Out of 2815

longitudinally ascertained subjects 538 subjects were also younger relatives of 1059 EA subjects from 118 extended families. There were around 726 subjects in these EA families that were not genotyped using the exome array. All of EA subjects from 118 families were previously genotyped using Illumina Human OmniExpress array 12.VI (Illumina, San Diego, CA, USA). This gave us an opportunity to infer the dense SNPs in un-genotyped subjects using identity by descent information generated through the sparse array using publicly available long range phasing program ChromoPhase (2). We phased genotyped subjects in each pedigree for each chromosome by combining the sparse and dense genotypes and used this IBD information to fill in the missing genotypes according to rules of Mendelian segregation. The phase of unambiguous SNPs were generated using the population frequency and were imputed according to population based imputation. Using this option we were able to guess > 98% missing haplotypes in all of the subjects. After infer process we removed the variants that didn't follow the rules of Mendelian segregation.

**One member of all pairs of related individuals between first UKB release (150K) and second UKB release (350K) were removed.

Table S2. Per-study, per-phenotype sample size and genomic control (European ancestry only).

	Cigarettes per Day		Pack Years		Age of Initiation of Smoking		Smoking Initiation		Drinks per Week	
Study	N	GC	N	GC	N	GC	N	GC	N	GC
ARIC	5381	1.063	5304	1.045	5407	1.096	8970	1.064	5966	1.000
COGA	1465	.895	1435	1.050	1638	0.923	-	-	3398	0.953
FTC	819	1.048	767	1.012	769	1.059	1467	1.063	1242	0.995
FUSION	568	1.040	530	1.042	562	1.018	1153	1.016	830	0.997
GECCO	2916	1.018	2876	1.028	-	-	6459	0.993	2077	0.967
HRS	3303	0.988	3303	0.992	3303	0.998	6393	1.096	4507	0.988
ID1000	366	0.974	373	1.007	-	-	803	0.994	740	0.985
MEC	1087	0.979	1082	0.963	1086	0.999	1903	0.973	1271	1.064
METSIM	1374	1.028	1370	1.016	1370	1.026	8146	1.044	6303	1.099
MHI	4391	1.011	4400	1.016	4420	1.018	6820	1.025	4205	1.022
MCTFR	2043	0.991	-	-	-	-	-	-	4757	0.998
NAGOZALC	671	1.006	646	1.006	647	1.011	1038	1.004	663	0.994
NESCOG	217	1.004	220	1.000	-	-	486	1.038	437	0.980
SardiNIA	1969	1.009	1967	1.064	1967	1.014	5069	1.082	2516	1.142
TwinsUK	358	1.039	358	1.010	358	1.006	878	0.971	603	0.989
WHI	6246	1.031	6236	1.006	-	-	-	-	7213	0.982
UK Biobank (MAF>1%)	120,744	1.10	120,126	1.08	124,590	1.03	383,631	1.15	311,126	1.06
UK Biobank (MAF≤1%)	[a]	1.03	[a]	1.01	[a]	.96	[a]	.98	[a]	1.02

Note: Study abbreviations are defined in Table S1.

[a]Sample sizes are the same for UK Biobank common and rare variants.

Table S3. Per-study, per-phenotype sample size and genomic control (African ancestry only).

	Cigarettes per Day		Pack Years		Age of Initiation of Smoking		Smoking Initiation		Drinks per Week	
Study	N	GC	N	GC	N	GC	N	GC	N	GC
COGA	476	0.93	457	0.99	494	0.91	-	-	1,182	0.94
HRS	961	1.03	961	1.02	961	1.01	1,746	1.03	980	0.99
UK Biobank (MAF>1%)	1,248	1.04	1,240	1.01	1,250	1.01	7,228	0.99	5432	1.04

Note: Study abbreviations are defined in Table S1.

Tables S4-S7 are available in Excel spreadsheets for convenience. See Supplement 2.

Table S8. Partition of heritablity for variants on exome array. We estimate the "chip" heritability for variants on the exome array using LD Score Regression. We consider a model that consists of seven functional categories. We report estimates of heritability ($\hat{h}^2$), their standard deviation $se(\hat{h}^2)$ as well as the p-value and z-score.

Annotation	$(\hat{h}^2)$	$se(\hat{h}^2)$	p-value	z-Score
Age of Initiation of Smoking				
(Intercept)	1	0.022	0	47
3' UTR	0.0046	0.0013	0.26	1.1
5' UTR	0.0089	0.0019	0.14	1.5
Common Coding Variants	0.014	0.0016	0.0069	2.7
Intergenic	0.016	0.0036	0.15	1.4
Intron	0.0042	0.00072	0.066	1.8
Rare Coding	0.011	0.0015	0.028	2.2
Synonymous	0.0017	$7.00´ 10^{-4}$	0.44	0.77
Cigarettes per Day				
(Intercept)	1	0.023	0	43
3' UTR	0.0044	0.00049	$1.90´ 10^{-1}$	1.3
5' UTR	0.0061	0.00072	$2.10´ 10^{-1}$	1.2
Common Coding Variants	0.025	$6.00´ 10^{-4}$	$1.70´ 10^{-9}$	6
Intergenic	0.027	0.0014	$3.10´ 10^{-3}$	3
Intron	0.0022	0.00027	$2.30´ 10^{-1}$	1.2
Rare Coding	0.0098	$6.00´ 10^{-4}$	$1.70´ 10^{-2}$	2.4
Synonymous	0.015	0.00058	$1.20´ 10^{-4}$	3.8
Drinks per Week				
(Intercept)	1.1	0.027	0	42
3' UTR	0.015	0.0023	$3.00´ 10^{-2}$	2.2
5' UTR	0.0095	0.0034	$3.60´ 10^{-1}$	0.92
Common Coding Variants	0.035	0.0029	$5.10´ 10^{-5}$	4.1
Intergenic	0.059	0.0065	$2.40´ 10^{-3}$	3
Intron	0.0042	0.0013	$2.80´ 10^{-1}$	1.1
Rare Coding	0.02	0.0013	$1.80´ 10^{-7}$	5.2
Synonymous	0.017	0.0028	$4.30´ 10^{-2}$	2

Annotation	$(\widehat{h}^2)$	$se(\widehat{h}^2)$	p-value	z-Score
Pack Years				
(Intercept)	1	0.024	0	43
3' UTR	0.0041	0.00056	$2.10' 10^{-1}$	1.3
5' UTR	0.0075	0.00082	$1.20' 10^{-1}$	1.6
Common Coding Variants	0.018	0.00069	$8.80' 10^{-6}$	4.4
Intergenic	0.038	0.0016	$3.70' 10^{-5}$	4.1
Intron	0.002	0.00031	$2.70' 10^{-1}$	1.1
Rare Coding	0.018	0.00068	$8.50' 10^{-6}$	4.5
Synonymous	0.012	0.00066	$1.30' 10^{-3}$	3.2
Smoking Initiation				
(Intercept)	1	$3.40' 10^{-2}$	$5.70' 10^{-206}$	31
3' UTR	0.019	$1.90' 10^{-4}$	$1.40' 10^{-17}$	8.5
5' UTR	0.04	$2.80' 10^{-4}$	$1.10' 10^{-33}$	12
Common Coding Variants	0.019	$2.30' 10^{-4}$	$5.40' 10^{-12}$	6.9
Intergenic	0.038	$5.30' 10^{-4}$	$6.20' 10^{-10}$	6.2
Intron	0.0024	$1.10' 10^{-4}$	$5.40' 10^{-2}$	1.9
Rare Coding	0.022	$2.30' 10^{-4}$	$3.90' 10^{-16}$	8.1
Synonymous	0.00025	$2.40' 10^{-4}$	$9.30' 10^{-1}$	0.09

Supplemental References

1. David, S.P., et al., *Genome-wide meta-analyses of smoking behaviors in African Americans.* Translational Psychiatry, 2012. **2**.
2. Daetwyler, H.D., et al., *Imputation of missing genotypes from sparse to high density using long-range phasing.* Genetics, 2011, **1**.

Supplementary Table 4. A comparison of our results with previous alcohol and nicotine focused, targeted resequencing based, genetic association studies.

			Original Study				GSCAN			
Phenotype	**rsID**	**Gene**	**Aggregate P-Value**	**Single P-Value**	**Study**	**Finding from Original Study**	**Pheno**	**VT P-Value**	**Single P-value**	**MAF**
Alcohol Dependence	rs115360541	SERINC2	-	0.005	Zuo, Wang et al., 2013	Replicated in study	Drinks per Week	0.038	-	-
Alcohol Dependence	-	ALDH2	-	-	Eng et al., 2007	Significant burden test	Drinks per Week	0.38	-	-
Alcohol Dependence	-	ADH1B	-	-	Eng et al., 2007	Significant burden test	Drinks per Week	0.31	-	-
Alcohol Dependence	-	ADH1C	-	-	Eng et al., 2007	Significant burden test	Drinks per Week	**6.00E-27**	-	-
Alcohol Dependence	rs149775276	CHRNB3	5.00E-04	2.60E-04	Haller, Kapoor et al., 2014	Top SNP in gene	Drinks per Week	0.45	0.97	0.0012
Alcohol Dependence	rs111797757	ADH1A	-	0.01	Peng et al., 2014	Top SNP in gene	Drinks per Week	0.26	0.078	0.087
Alcohol Dependence	rs12507078	ADH6	-	0.003	Peng et al., 2014	Top SNP in gene	Drinks per Week	0.6	0.32	0.083
Alcohol Dependence	rs145341314	ADH5/4	-	0.003	Peng et al., 2014	Top SNP in gene	Drinks per Week	0.72	0.7	0.081
Alcohol Dependence	rs1497372	ADH1C	-	0.004	Peng et al., 2014	Top SNP in gene	Drinks per Week	**6.00E-27**	-	-
Alcohol Dependence	rs17588403	ADH7	-	0.03	Peng et al., 2014	Top SNP in gene	Drinks per Week	0.88	0.12	0.19
Alcohol Dependence	rs190914158	ALDH2	-	0.009	Peng et al., 2014	Top SNP in gene	Drinks per Week	0.38	-	-
Alcohol Dependence	rs2226896	ADH4/6	-	0.003	Peng et al., 2014	Top SNP in gene	Drinks per Week	0.1	0.33	0.082
Alcohol Dependence	rs28914770	ADH1B	-	0.018	Peng et al., 2014	Top SNP in gene	Drinks per Week	0.31	0.075	0.087
Alcohol Dependence	rs7375388	ADH6/1A	-	0.003	Peng et al., 2014		Drinks per Week	0.6	1	0.086
Alcohol Dependence	rs1229984	ADH1B	-	5.88E-05	Way et al., 2015	Top SNP in gene	Drinks per Week	0.31	**2.27E-173**	0.02
Alcohol Dependence	rs1789891	ADH1B/ADH1(	-	5.31E-05	Way et al., 2015	Top SNP in gene	Drinks per Week	0.31	**9.14E-19**	0.18
Alcohol Dependence	rs35961897	SERINC2	1.60E-04	4.10E-05	Zuo, Wang et al., 2013	Top SNP in gene	Drinks per Week	0.038	0.41	0.048
Alcohol Dependence	rs16834507	SERINC2	-	0.01	Zuo, Wang et al., 2013	Top SNP in gene	Drinks per Week	0.038	0.086	6.12E-05
Alcohol Dependence	rs77840364	SERINC2	-	0.02	Zuo, Wang et al., 2013	Top SNP in gene	Drinks per Week	0.038	-	-
Alcohol Dependence	rs79051763	PTP4A1	4.20E-03	0.006	Zuo, Wang et al., 2013	Top SNP in gene	Drinks per Week	-	-	-
Alcohol Dependence	rs114282789	EYS	0.23	0.02	Zuo, Wang et al., 2013	Top SNP in gene	Drinks per Week	0.36	-	-
Alcohol Dependence	rs319919	EYS	0.34	9.50E-04	Zuo, Wang et al., 2013	Top SNP in gene	Drinks per Week	0.36	0.21	0.29
Nicotine Dependence	-	CHRNB4	6.00E-05	-	Haller, Druley et al., 2012	Significant burden test	Cigarettes per Day	0.24	-	-
Nicotine Dependence	-	CHRNA4	0.04	-	Wessel et al., 2010	Significant burden test	Cigarettes per Day	0.24	-	-
Nicotine Dependence	-	DBH	1.00E-06	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.84	-	-
Nicotine Dependence	-	NRXN3	1.00E-06	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.4	-	-
Nicotine Dependence	-	NRXN1	2.00E-06	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.42	-	-
Nicotine Dependence	-	TAS2R38	2.00E-06	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.22	-	-
Nicotine Dependence	-	CHRNA9	8.00E-06	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.49	-	-
Nicotine Dependence	-	GRIN3A	8.00E-06	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.96	-	-
Nicotine Dependence	-	CDH13	3.50E-05	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.18	-	-
Nicotine Dependence	-	ARRB2	1.32E-04	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.43	-	-
Nicotine Dependence	-	DNM1	3.53E-04	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.37	-	-
Nicotine Dependence	-	NTRK2	4.25E-04	-	Yang et al., 2015	Significant burden test	Cigarettes per Day	0.78	-	-
Nicotine Dependence	-	CHRNA4	1.90E-39	-	Zuo et al., 2016	Significant burden test	Cigarettes per Day	0.12	-	-
Nicotine Dependence	-	CHRNA9	6.10E-30	-	Zuo et al., 2016	Significant burden test	Cigarettes per Day	0.49	-	-
Nicotine Dependence	-	CHRNA10	3.40E-29	-	Zuo et al., 2016	Significant burden test	Cigarettes per Day	0.47	-	-
Nicotine Dependence	-	CHRNA7	6.10E-27	-	Zuo et al., 2016	Significant burden test	Cigarettes per Day	0.91	-	-
Nicotine Dependence	-	CHRNB2	1.30E-18	-	Zuo et al., 2016	Significant burden test	Cigarettes per Day	0.2	-	-
Nicotine Dependence	-	CHRNE	4.90E-15	-	Zuo et al., 2016	Significant burden test	Cigarettes per Day	0.57	-	-
Nicotine Dependence	-	CHRNA1	1.20E-14	-	Zuo et al., 2016	Significant burden test	Cigarettes per Day	0.54	-	-
Nicotine Dependence	-	CHRNA2	5.20E-13	-	Zuo et al., 2016	Significant burden test	Cigarettes per Day	0.48	-	-
Nicotine Dependence	rs16969968	CHRNA5	0.01	0.003	Olfson et al., 2016	Top SNP in gene	Cigarettes per Day	0.19	**2.52E-139**	0.34

Nicotine Dependence	rs2229961	CHRNA5	-	0.03	Olfson et al., 2016	Top SNP in gene	Cigarettes per Day	0.19	**2.01E-14**	0.019
Nicotine Dependence	rs56175056	CHRNA4	-	1.20E-04	Thorgeirsson et al., 2016	Top SNP in gene	Cigarettes per Day	0.12	0.11	2.67E-04
Nicotine Dependence	rs2072661	CHRNB2	2.00E-03	0.002	Wessel et al., 2010	Top SNP in gene	Cigarettes per Day	0.2	0.59	0.24

Supplementary Table 5. Sentinel GWAS Variants

PHENOTYPE	CHROM	POS	rsID	REF	ALT	AF	STAT	PVALUE	BETA	SE	N	Annotation [Nearest Gene]	Distance to Gene
Age of Initiation of Smoking	3	85682087	rs12493563	G	T	0.341	43	5.61×10^{-11}	0.0277	0.00422	124590	Intron[CADM2]	0
Age of Initiation of Smoking	17	31539143	rs8082191	A	T	0.279	32.4	1.27×10^{-8}	0.0254	0.00447	124590	Intron[ACCN1]	0
Age of Initiation of Smoking	20	14844018	rs442467	T	C	0.676	42.5	7.18×10^{-11}	0.0279	0.00428	124590	Intron[MACROD2]	0
Cigarettes per Day	1	154548521	rs2072659	C	G	0.102	30	4.28×10^{-8}	-0.0369	0.00673	120744	Utr3[CHRNB2]	0
Cigarettes per Day	4	2961713	rs2488808	T	A	0.308	31.1	2.51×10^{-8}	-0.0246	0.00441	120744	Intron[NOP14]	0
Cigarettes per Day	7	32273107	rs7806224	T	C	0.371	41.9	9.51×10^{-11}	0.0244	0.00377	150484	Intron[PDE1C]	0
Cigarettes per Day	8	42550498	rs6474412	C	T	0.775	66.6	3.40×10^{-16}	0.0352	0.00432	153918	Intergenic[CHRNB3]	2032
Cigarettes per Day	8	64567670	rs1217106	A	G	0.782	34.9	3.51×10^{-9}	-0.0291	0.00493	120744	Intergenic[YTHDF3]	442388
Cigarettes per Day	9	136478355	rs3025343	G	A	0.117	43	5.54×10^{-11}	0.0368	0.00561	153918	Intergenic[DBH]	23154
Cigarettes per Day	12	4268731	rs113469106	G	A	0.057	33.1	8.74×10^{-9}	0.0505	0.00877	120744	Intergenic[CCND2]	114129
Cigarettes per Day	15	78806023	rs8034191	T	C	0.341	633	1.41×10^{-139}	0.0956	0.0038	153918	Intron[AGPHD1]	0
Cigarettes per Day	17	37814687	rs36015615	G	A	0.000163	30.6	3.18×10^{-8}	1.16	0.20948	69951	Nonsynonymous[STARD3]	0
Cigarettes per Day	19	41353107	rs56113850	T	C	0.575	305	2.99×10^{-68}	0.0719	0.00412	120744	Intron[CYP2A6]	0
Cigarettes per Day	20	61984317	rs6011779	C	T	0.806	37.9	7.43×10^{-10}	-0.0317	0.00514	120744	Intron[CHRNA4]	0
Pack Years	1	110045880	rs4370783	A	C	0.115	30.7	3.06×10^{-8}	-0.0321	0.00579	146945	Intergenic[CYB561D1]	2823
Pack Years	2	45154418	rs7569203	A	C	0.317	30.4	3.57×10^{-8}	0.0218	0.00396	146945	Intergenic[SIX3]	14915
Pack Years	2	105969362	rs62155873	C	T	0.127	36.2	1.75×10^{-9}	0.0334	0.00555	146945	Intergenic[C2orf49]	7568
Pack Years	7	32273107	rs7806224	T	C	0.375	67.1	2.57×10^{-16}	0.0287	0.0035	174410	Intron[PDE1C]	0
Pack Years	8	42550498	rs6474412	C	T	0.77	64	1.24×10^{-15}	0.0319	0.00398	177812	Intergenic[CHRNB3]	2032
Pack Years	9	136478355	rs3025343	G	A	0.114	90.1	2.28×10^{-21}	0.0501	0.00527	177812	Intergenic[DBH]	23214
Pack Years	11	43603300	rs11820132	T	C	0.44	36.3	1.70×10^{-9}	-0.0224	0.00372	146945	Intergenic[MIR670]	22013
Pack Years	13	112179953	rs2026174	C	T	0.504	37.8	7.94×10^{-10}	0.0227	0.00369	146945	Intergenic[C13orf16]	183385
Pack Years	14	104146421	rs4900590	C	T	0.324	35.4	2.73×10^{-9}	0.0234	0.00394	146945	Intron[KLC1]	0
Pack Years	15	78806023	rs8034191	T	C	0.338	533	5.81×10^{-118}	0.0819	0.00355	177812	Intron[AGPHD1]	0
Pack Years	15	89928189	rs150353	T	G	0.448	30.8	2.91×10^{-8}	-0.0205	0.0037	147591	Intron[LOC254559]	0
Pack Years	16	29311500	rs536748263	C	T	0.00126	31	2.60×10^{-8}	0.29	0.0521	146945	Intron[RUNDC2C]	0
Pack Years	16	69972486	rs9302605	T	A	0.771	34	5.62×10^{-9}	0.0256	0.00439	146945	Intron[WWP2]	0
Pack Years	16	89645437	rs369230	G	T	0.695	35	3.26×10^{-9}	0.0237	0.004	146945	Intron[CPNE7]	0
Pack Years	17	27447905	rs12601994	C	T	0.156	32.1	1.50×10^{-8}	0.0288	0.00509	146945	Intron[MYO18A]	0
Pack Years	19	41353107	rs56113850	T	C	0.572	107	5.61×10^{-25}	0.0385	0.00373	146945	Intron[CYP2A6]	0
Pack Years	20	61991833	rs45497800	C	T	0.0867	68.2	1.50×10^{-16}	0.0541	0.00655	146945	Intron[CHRNA4]	0
Smoking Initiation	1	44076019	rs11210887	G	A	0.7	48.7	3.00×10^{-12}	-0.0174	0.00249	384669	Intron[PTPRF]	0
Smoking Initiation	1	66470206	rs2186122	A	T	0.561	43.5	4.25×10^{-11}	0.0152	0.0023	383631	Intron[PDE4B]	0
Smoking Initiation	1	73882478	rs1475064	A	G	0.615	30	4.38×10^{-8}	-0.0121	0.00221	433216	Intergenic[LRRIQ3]	609322
Smoking Initiation	1	74991644	rs1514175	A	G	0.575	40.4	2.07×10^{-10}	-0.0138	0.00217	433216	Intron[FPGT-TNNI3K|TNNI3I	0
Smoking Initiation	1	91191582	rs72720396	A	G	0.231	38.1	6.85×10^{-10}	-0.0167	0.00271	383631	Intergenic[BARHL2]	8954
Smoking Initiation	1	154154194	rs35761479	G	A	0.122	52.5	4.26×10^{-13}	-0.0253	0.00349	383631	Intron[TPM3]	0
Smoking Initiation	1	210359333	rs55921136	T	C	0.203	31	2.56×10^{-8}	-0.0158	0.00284	383631	Intergenic[SYT14]	22007
Smoking Initiation	2	615140	rs7567570	T	C	0.827	46.5	8.96×10^{-12}	0.0206	0.00302	383631	Intergenic[TMEM18]	52827
Smoking Initiation	2	45154908	rs528301	G	A	0.552	51.6	6.93×10^{-13}	0.0165	0.00229	384669	Intergenic[SIX3]	14387
Smoking Initiation	2	60024857	rs7585579	C	G	0.489	47.4	5.65×10^{-12}	0.0157	0.00228	383631	Intergenic[BCL11A]	653492
Smoking Initiation	2	60526747	rs1029984	G	T	0.563	32.1	1.45×10^{-8}	0.013	0.0023	383631	Intergenic[BCL11A]	151602

Smoking Initiation	2	64092819	rs140144265	C	T	0.00232	29.7	4.92×10^{-8}	-0.129	0.0237	383631	Intron[UGP2]	0
Smoking Initiation	2	104150791	rs6720941	C	T	0.461	65.4	6.06×10^{-16}	0.0185	0.00229	383631	Intergenic[TMEM182]	716605
Smoking Initiation	2	137542847	rs35702515	G	T	0.237	34.3	4.84×10^{-9}	0.0157	0.00269	383631	Intergenic[THSD7B]	205589
Smoking Initiation	2	145400317	rs1427499	A	G	0.711	40	2.57×10^{-10}	-0.0159	0.00252	383631	Intergenic[DKFZp686O1327]	25192
Smoking Initiation	2	146155371	rs6745444	A	G	0.482	85.5	2.36×10^{-20}	0.0211	0.00228	383631	Intergenic[DKFZp686O1327]	321129
Smoking Initiation	2	155742894	rs2652434	T	C	0.533	31.4	2.09×10^{-8}	0.0128	0.00229	383631	Intergenic[KCNJ3]	30043
Smoking Initiation	2	162891848	rs6432708	C	T	0.581	49.1	2.38×10^{-12}	-0.0162	0.00231	383631	Intron[DPP4]	0
Smoking Initiation	3	5723818	rs4479577	C	T	0.482	38.7	4.85×10^{-10}	0.0142	0.00228	383631	Intergenic[EDEM1]	462139
Smoking Initiation	3	25193102	rs904592	C	T	0.482	30.2	3.80×10^{-8}	0.0126	0.00228	383631	Intergenic[RARB]	276793
Smoking Initiation	3	52874288	rs6445538	T	C	0.233	42.4	7.58×10^{-11}	0.0165	0.00254	433216	Utr3[TMEM110]	0
Smoking Initiation	3	61819535	rs35995169	A	G	0.402	33.3	8.00×10^{-9}	0.0134	0.00233	383631	Intron[PTPRG]	0
Smoking Initiation	3	85460131	rs1549979	C	T	0.622	71.3	3.12×10^{-17}	-0.0187	0.00222	433216	Intron[CADM2]	0
Smoking Initiation	3	86149109	rs73138150	A	T	0.324	39	4.20E-10	-0.0152	0.00244	383631	Intergenic[CADM2]	25613
Smoking Initiation	3	117639575	rs62264764	G	A	0.149	37.5	9.05×10^{-10}	-0.0196	0.0032	383631	Intergenic[IGSF11]	979874
Smoking Initiation	3	118302515	rs12053870	T	G	0.545	33.3	7.95×10^{-9}	0.0132	0.00229	383631	Intergenic[IGSF11]	316934
Smoking Initiation	3	157393770	rs963354	C	A	0.672	33.5	6.98×10^{-9}	0.0141	0.00243	384669	Intergenic[C3orf55]	74704
Smoking Initiation	4	28540930	rs13125329	A	C	0.258	31.9	1.63×10^{-8}	0.0147	0.00261	383631	Intergenic[MIR4275]	280254
Smoking Initiation	4	57740334	rs6835108	A	G	0.215	36	1.94×10^{-9}	-0.0167	0.00278	383631	Intergenic[REST]	33707
Smoking Initiation	4	143617304	rs11943397	T	C	0.628	35.8	2.22×10^{-9}	0.0141	0.00236	383631	Intron[INPP4B]	0
Smoking Initiation	4	147948150	rs3827592	G	A	0.351	40.1	2.46×10^{-10}	-0.0151	0.00239	383631	Intergenic[TTC29]	81127
Smoking Initiation	5	22219503	rs4518351	A	G	0.444	32.4	1.28×10^{-8}	0.0131	0.00229	384669	Intron[CDH12]	0
Smoking Initiation	5	60121470	rs62372074	A	T	0.148	34.7	3.82×10^{-9}	0.019	0.00322	383631	Intron[ELOVL7]	0
Smoking Initiation	5	80263403	rs13357015	G	A	0.635	30.4	3.54×10^{-8}	0.0131	0.00237	383631	Intron[RASGRF2]	0
Smoking Initiation	5	87781168	rs10044618	C	T	0.432	47.5	5.39×10^{-12}	-0.0159	0.0023	383631	Intergenic[LOC645323]	55488
Smoking Initiation	5	94202167	rs27003	T	C	0.694	33.3	8.02×10^{-9}	0.0143	0.00248	383631	Intron[MCTP1]	0
Smoking Initiation	5	106825618	rs72789626	T	A	0.138	35.7	2.33×10^{-9}	-0.0198	0.00331	383631	Intron[EFNA5]	0
Smoking Initiation	5	157743295	rs11135030	C	T	0.306	37.7	8.17×10^{-10}	0.0152	0.00248	383631	Intergenic[EBF1]	379662
Smoking Initiation	5	166996722	rs1549212	C	T	0.626	37.1	1.14×10^{-9}	-0.0144	0.00236	383631	Intron[ODZ2]	0
Smoking Initiation	6	28366151	rs2232423	A	G	0.111	32.1	1.43×10^{-8}	-0.0194	0.00342	433216	Nonsynonymous[ZSCAN12]	0
Smoking Initiation	6	29548089	rs926552	G	A	0.134	32.8	1.00×10^{-8}	-0.0181	0.00315	433216	Intergenic[SNORD32B]	1951
Smoking Initiation	6	30857894	rs1264322	G	A	0.146	31.7	1.77×10^{-8}	-0.0171	0.00304	433216	Intron[DDR1]	0
Smoking Initiation	6	31686497	rs15574	G	A	0.194	33.5	7.12×10^{-9}	0.0159	0.00275	422914	Utr3[LY6G6C]	0
Smoking Initiation	6	48121843	rs59263024	T	TA	0.1	30.5	3.26×10^{-8}	0.0388	0.00701	112811	Intergenic[C6orf138]	85379
Smoking Initiation	6	67546542	rs12205538	T	C	0.387	36.4	1.60×10^{-9}	0.0141	0.00234	383631	Intergenic[MCART3P]	1047285
Smoking Initiation	6	93833436	rs6938042	T	G	0.339	36	2.01×10^{-9}	0.0144	0.00241	384669	Intergenic[EPHA7]	116270
Smoking Initiation	6	98888740	rs2132029	C	T	0.318	39.5	3.20×10^{-10}	0.0154	0.00245	383631	Intergenic[POU3F2]	393926
Smoking Initiation	6	101153907	rs6922219	G	C	0.47	31.2	2.28×10^{-8}	0.0128	0.00229	383631	Intron[ASCC3]	0
Smoking Initiation	6	111644332	rs240963	T	C	0.841	72.1	2.03×10^{-17}	-0.0265	0.00312	383631	Intron[REV3L]	0
Smoking Initiation	7	1701592	rs10255004	C	T	0.432	42.2	8.28×10^{-11}	-0.015	0.0023	383631	Intergenic[TFAMP1]	45261
Smoking Initiation	7	3448973	rs2056475	A	G	0.746	42.9	5.65×10^{-11}	-0.0172	0.00262	383631	Intron[SDK1]	0
Smoking Initiation	7	96624257	rs2240294	T	A	0.444	30	4.25×10^{-8}	-0.0126	0.0023	383631	Intron[DLX6-AS1]	0
Smoking Initiation	7	117523709	rs10233018	A	G	0.504	43	5.53×10^{-11}	0.015	0.00228	383631	Intergenic[CTTNBP2]	10264
Smoking Initiation	8	10153082	rs17151637	C	T	0.281	35.9	2.12×10^{-9}	-0.0152	0.00254	383631	Intron[MSRA]	0
Smoking Initiation	8	10836472	rs7010246	G	A	0.365	30.7	3.04×10^{-8}	-0.0131	0.00237	383631	Intron[XKR6]	0

Smoking Initiation	8	27426077	rs1565735	T	A	0.201	67.8	1.77×10^{-16}	-0.0235	0.00285	383631	Intergenic[EPHX2]	23546
Smoking Initiation	8	59817068	rs1562612	A	G	0.504	37.4	9.73×10^{-10}	-0.0139	0.00228	384669	Intron[TOX]	0
Smoking Initiation	8	92775372	rs10956808	T	G	0.422	36.1	1.86×10^{-9}	-0.0131	0.00218	433216	Intergenic[RUNX1T1]	191934
Smoking Initiation	9	16758107	rs10962568	G	C	0.149	38.7	5.00×10^{-10}	0.0199	0.00321	383631	Intron[BNC2]	0
Smoking Initiation	9	86761781	rs1246264	T	C	0.695	38.8	4.82×10^{-10}	0.0154	0.00248	384669	Intergenic[SLC28A3]	128968
Smoking Initiation	9	128134034	rs7870475	T	C	0.475	44.4	2.73×10^{-11}	0.0152	0.00229	383631	Intergenic[GAPVD1]	6778
Smoking Initiation	9	137975748	rs7019640	C	T	0.23	32.7	1.10×10^{-8}	0.0155	0.00271	383631	Intron[OLFM1]	0
Smoking Initiation	10	21783634	rs12770228	G	A	0.315	42.2	8.24×10^{-11}	0.016	0.00246	383631	Utr3[C10orf114]	0
Smoking Initiation	10	27397454	rs11015535	C	T	0.093	31.1	2.46×10^{-8}	0.0404	0.00725	112811	Intergenic[YME1L1]	1966
Smoking Initiation	10	63674885	rs7921378	G	C	0.481	42.8	6.06×10^{-11}	-0.0149	0.00228	383631	Intron[ARID5B]	0
Smoking Initiation	10	104045951	rs375915173	GT	G	0.828	31.7	1.75×10^{-8}	-0.0314	0.00557	112811	Deletion[GBF1]	0
Smoking Initiation	10	104643241	rs12783782	G	A	0.257	59.9	1.00×10^{-14}	0.0202	0.00261	383631	Intron[AS3MT|C10orf32-AS3	0
Smoking Initiation	10	125680419	rs9423279	C	G	0.657	33.3	7.90×10^{-9}	-0.0139	0.0024	383631	Intergenic[CPXM2]	28954
Smoking Initiation	11	7950797	rs4523689	A	G	0.393	32.4	1.25×10^{-8}	-0.0133	0.00234	383631	Intergenic[OR10A6]	620
Smoking Initiation	11	20129311	rs35891966	G	A	0.0719	32.4	1.27×10^{-8}	-0.0237	0.00416	433216	Nonsynonymous[NAV2]	0
Smoking Initiation	11	27694241	rs2049045	G	C	0.187	46.4	9.77×10^{-12}	-0.0199	0.00293	383631	Intron[BDNF|BDNF-AS1]	0
Smoking Initiation	11	85927213	rs1466802	T	C	0.775	38.1	6.68×10^{-10}	0.0169	0.00274	383631	Intergenic[EED]	28562
Smoking Initiation	11	112839532	rs7948789	A	G	0.386	116	6.11×10^{-27}	0.0252	0.00234	383631	Intron[NCAM1]	0
Smoking Initiation	12	56412487	rs1701704	T	G	0.341	30.9	2.70×10^{-8}	-0.0126	0.00227	433216	Intergenic[IKZF4]	2160
Smoking Initiation	12	57976118	rs11172256	A	G	0.257	29.8	4.80×10^{-8}	0.0143	0.00261	383631	Intron[KIF5A]	0
Smoking Initiation	12	69686203	rs317660	G	A	0.718	35.8	2.23×10^{-9}	-0.0152	0.00254	383631	Intergenic[CPSF6]	18135
Smoking Initiation	12	111904371	rs4766578	T	A	0.504	30.8	2.83×10^{-8}	-0.0127	0.00228	384669	Intron[ATXN2]	0
Smoking Initiation	12	121111214	rs34067374	C	T	0.14	30.8	2.80×10^{-8}	0.0183	0.00329	383631	Intergenic[CABP1]	6058
Smoking Initiation	12	133451277	rs80032242	G	A	0.137	30.4	3.54×10^{-8}	0.0183	0.00332	383631	Intron[CHFR]	0
Smoking Initiation	13	38359684	rs4460944	T	C	0.527	31.8	1.68×10^{-8}	0.0129	0.00229	383631	Intron[TRPC4]	0
Smoking Initiation	13	100548329	rs7322872	C	T	0.791	34	5.55×10^{-9}	-0.0164	0.00281	383631	Utr3[CLYBL]	0
Smoking Initiation	13	101179012	rs837333	T	C	0.477	35.7	2.24×10^{-9}	0.0137	0.00229	383631	Intron[PCCA]	0
Smoking Initiation	14	29500106	rs7141236	T	C	0.187	34.6	4.01×10^{-9}	-0.0172	0.00292	384669	Intergenic[C14orf23]	236110
Smoking Initiation	14	90030280	rs171460	G	A	0.26	31	2.52×10^{-8}	-0.0145	0.0026	383631	Intron[FOXN3]	0
Smoking Initiation	15	47935843	rs1435741	G	A	0.43	55	1.18×10^{-13}	0.0171	0.00231	383631	Intron[SEMA6D]	0
Smoking Initiation	15	54119670	rs766614	T	G	0.212	31.8	1.75×10^{-8}	0.0157	0.00279	384669	Intergenic[WDR72]	67762
Smoking Initiation	16	735921	rs763053	T	C	0.225	34.3	4.70×10^{-9}	-0.016	0.00273	384669	Synonymous[WDR24]	0
Smoking Initiation	16	18058548	rs77878475	T	A	0.0858	39.5	3.31×10^{-10}	-0.0256	0.00408	383631	Intergenic[NOMO2]	455502
Smoking Initiation	16	69567781	rs11646575	G	A	0.439	39.6	3.13×10^{-10}	0.0145	0.0023	383631	Intergenic[NFAT5]	31306
Smoking Initiation	18	77458644	rs522180	C	T	0.353	45.6	1.47×10^{-11}	0.0161	0.00239	383631	Intron[CTDP1]	0
Smoking Initiation	19	4474725	rs76608582	C	A	0.0482	32.6	1.15×10^{-8}	-0.0304	0.00533	383631	Intron[HDGFRP2]	0
Smoking Initiation	20	29893501	rs6061162	C	T	0.912	29.8	4.83×10^{-8}	0.022	0.00402	383631	Intron[DEFB116]	0
Smoking Initiation	20	31140165	rs4911241	C	T	0.239	51.9	5.88×10^{-13}	0.0182	0.00252	433216	Intergenic[LOC149950]	35093
Drinks per Week	1	4569436	rs780569	T	A	0.708	33.2	8.47×10^{-9}	-0.0161	0.00279	311126	Intergenic[LOC284661]	84866
Drinks per Week	1	95730588	rs3886141	G	C	0.409	31.5	1.96×10^{-8}	-0.0145	0.00258	311126	Intergenic[RWDD3]	17791
Drinks per Week	1	165119792	rs10753661	G	A	0.684	33.2	8.36×10^{-9}	-0.0157	0.00273	311126	Intergenic[LMX1A]	51267
Drinks per Week	2	27730940	rs1260326	T	C	0.608	176	4.58×10^{-40}	0.0321	0.00242	357854	Nonsynonymous[GCKR]	0
Drinks per Week	2	44271496	rs75120545	C	T	0.0302	33.7	6.29×10^{-9}	-0.043	0.00741	311126	Intergenic[LRPPRC]	48316
Drinks per Week	2	45154908	rs528301	G	A	0.553	69.6	7.12×10^{-17}	0.0213	0.00255	311789	Intergenic[SIX3]	14387

Drinks per Week	2	144272376	rs4233567	C	T	0.364	37.5	8.92×10^{-10}	-0.0161	0.00263	311126	Intron[ARHGAP15]	0
Drinks per Week	3	85049088	rs1872552	G	A	0.399	44.4	2.73×10^{-11}	0.0172	0.00259	311789	Intron[CADM2]	0
Drinks per Week	3	85593584	rs6780346	C	T	0.621	41.1	1.47×10^{-10}	-0.0168	0.00261	311126	Intron[CADM2]	0
Drinks per Week	4	39413780	rs28712821	G	A	0.608	187	1.67×10^{-42}	0.0355	0.0026	311126	Intron[KLB]	0
Drinks per Week	4	42117559	rs16854020	G	A	0.13	37.3	1.02×10^{-9}	0.023	0.00377	311126	Utr3[BEND4]	0
Drinks per Week	4	97853880	rs143118954	T	A	0.00233	35.8	2.16×10^{-9}	-0.157	0.02627	311126	Intergenic[C4orf37]	626164
Drinks per Week	4	98373818	rs143475242	C	T	0.00341	49.5	2.00×10^{-12}	-0.153	0.02175	311126	Intergenic[C4orf37]	106226
Drinks per Week	4	98979582	rs1727284	C	G	0.00147	50.4	1.26×10^{-12}	-0.235	0.03305	311126	Intron[C4orf37]	0
Drinks per Week	4	99713350	rs144198753	C	T	0.00935	304	4.41×10^{-68}	-0.23	0.01317	311126	Intergenic[EIF4E]	86341
Drinks per Week	4	100239319	rs1229984	T	C	0.98	788	2.27×10^{-173}	0.246	0.00878	334588	Nonsynonymous[ADH1B]	0
Drinks per Week	4	100919254	rs150021439	C	T	0.0044	96.1	1.12×10^{-22}	-0.188	0.01914	311126	Intergenic[LOC256880]	45686
Drinks per Week	4	103188709	rs13107325	C	T	0.0721	70.3	4.97×10^{-17}	-0.0383	0.00457	357854	Nonsynonymous[SLC39A8]	0
Drinks per Week	4	143617304	rs11943397	T	C	0.629	32.1	1.43×10^{-8}	0.0149	0.00262	311126	Intron[INPP4B]	0
Drinks per Week	6	12903957	rs9349379	A	G	0.406	30.8	2.85×10^{-8}	0.0134	0.00241	357854	Intron[PHACTR1]	0
Drinks per Week	7	69793380	rs6978841	T	A	0.184	31.3	2.24×10^{-8}	-0.0183	0.00327	311126	Intron[AUTS2]	0
Drinks per Week	7	153489725	rs6969458	G	A	0.466	44.6	2.45×10^{-11}	0.017	0.00254	311126	Intergenic[DPP6]	94696
Drinks per Week	9	108755622	rs55932213	A	G	0.746	32.2	1.41×10^{-8}	0.0165	0.00291	311126	Intergenic[TMEM38B]	218228
Drinks per Week	9	109330236	rs4743009	G	A	0.201	33.5	6.98×10^{-9}	-0.0183	0.00317	311126	Intergenic[ZNF462]	295381
Drinks per Week	10	102626510	rs61873510	G	T	0.326	33.5	7.14×10^{-9}	-0.0157	0.0027	311126	Intergenic[PAX2]	36831
Drinks per Week	11	33204159	rs4756121	G	C	0.688	42.8	6.08×10^{-11}	0.0179	0.00274	311126	Intron[LOC338739]	0
Drinks per Week	11	47457539	rs11604680	A	G	0.32	40.8	1.69×10^{-10}	-0.0174	0.00272	311126	Intergenic[RAPSN]	1776
Drinks per Week	11	113412443	rs4309187	A	C	0.684	45.9	1.25×10^{-11}	0.0185	0.00273	311126	Intergenic[DRD2]	66456
Drinks per Week	11	116073667	rs1894071	A	G	0.857	30.2	3.94×10^{-8}	0.0199	0.00362	311126	Intergenic[LOC283143]	442792
Drinks per Week	11	121501406	rs10790449	T	C	0.465	33.6	6.91×10^{-9}	0.0147	0.00254	311126	Utr3[SORL1]	0
Drinks per Week	14	57281154	rs962961	C	T	0.326	34.7	3.88×10^{-9}	-0.0159	0.00271	311126	Intron[OTX2OS1]	0
Drinks per Week	14	58833909	rs57044214	G	A	0.284	35.1	3.17×10^{-9}	0.0166	0.00281	311126	Intron[ARID4A]	0
Drinks per Week	14	94844947	rs28929474	C	T	0.0198	44.8	2.22×10^{-11}	-0.0568	0.00849	357854	Nonsynonymous[SERPINA1]	0
Drinks per Week	16	24810681	rs17177078	C	T	0.0554	34.9	3.45×10^{-9}	-0.0327	0.00554	311126	Intron[TNRC6A]	0
Drinks per Week	16	28539848	rs4788084	C	T	0.418	31.1	2.46×10^{-8}	-0.0134	0.0024	355949	Intergenic[NUPR1]	8923
Drinks per Week	16	69745145	rs1800566	G	A	0.183	31.5	1.99×10^{-8}	0.0172	0.00306	357854	Nonsynonymous[NQO1]	0
Drinks per Week	16	72356964	rs11648570	T	C	0.107	31.5	2.04×10^{-8}	0.023	0.00409	311126	Intergenic[PMFBP1]	150733
Drinks per Week	16	73912588	rs1104608	G	C	0.426	40.6	1.87×10^{-10}	-0.0163	0.00256	311126	Intergenic[PSMD7]	418045
Drinks per Week	17	43754850	rs62055691	T	C	0.225	99.7	1.78×10^{-23}	-0.0303	0.00303	311126	Intron[CRHR1]	0
Drinks per Week	17	44285531	rs2469933	G	A	0.225	94.7	2.27×10^{-22}	-0.0295	0.00303	311789	Intron[KIAA1267]	0
Drinks per Week	17	44843136	rs199528	C	T	0.223	88.4	5.27×10^{-21}	-0.0286	0.00304	311126	Intron[WNT3]	0
Drinks per Week	18	40735531	rs62092069	G	A	0.373	31.9	1.60×10^{-8}	0.0148	0.00262	311126	Intergenic[RIT2]	39911
Drinks per Week	19	49248730	rs838145	G	A	0.541	46.6	8.86×10^{-12}	-0.0174	0.00254	311126	Intron[IZUMO1]	0
Drinks per Week	20	18693080	rs62217022	A	G	0.396	32.6	1.10×10^{-8}	-0.0148	0.00259	311126	Intron[DTD1]	0
Drinks per Week	20	31607551	rs61734341	G	C	0.0598	32.7	1.09×10^{-8}	-0.0285	0.00499	357854	Nonsynonymous[BPIL1]	0
Drinks per Week	22	41941243	rs9607812	G	A	0.187	38	7.21×10^{-10}	-0.02	0.00325	311126	Intergenic[POLR3H]	719
Drinks per Week	22	46481623	rs17884691	G	A	0.248	32.7	1.08×10^{-8}	-0.0168	0.00294	311126	Intergenic[LOC400931]	208

Supplementary Table 6. Bayesian Fine Mapping of all GWAS-significant variants, with Prior Informed by Statistical Association Only

PHENOTYPE	CHROM	POS	rsID	REF	ALT	AF	STAT	PVALUE	BETA	SE	N	SNPs in Interval	Interval Siz	Interval Start	Interval End	Annotation [Nearest Gene]
Age of Initiation of Smoking	3	85682087	rs1249356	G	T	0.341	43	5.61×10^{-11}	0.0277	0.00422	124590	59	177051	85521990	85699040	Intron[CADM2]
Age of Initiation of Smoking	17	31539143	rs8082191	A	T	0.279	32.4	1.27×10^{-8}	0.0254	0.00447	124590	50	32052	31539143	31571194	Intron[ACCN1]
Age of Initiation of Smoking	20	14844018	rs442467	T	C	0.676	42.5	7.18×10^{-11}	0.0279	0.00428	124590	29	67700	14817453	14885152	Intron[MACROD2]
Cigarettes per Day	1	154548521	rs2072659	C	G	0.102	30	4.28×10^{-8}	-0.0369	0.00673	120744	2	472	154548521	154548992	Utr3[CHRNB2]
Cigarettes per Day	4	2961713	rs2488808	T	A	0.308	31.1	2.51×10^{-8}	-0.0246	0.00441	120744	121	154665	2873835	3028499	Intron[NOP14]
Cigarettes per Day	7	32273107	rs7806224	T	C	0.371	41.9	9.51×10^{-11}	0.0244	0.00377	150484	48	117761	32261458	32379218	Intron[PDE1C]
Cigarettes per Day	8	42550498	rs6474412	C	T	0.775	66.6	3.40×10^{-16}	0.0352	0.00432	153918	1	1	42550498	42550498	Intergenic[CHRNB3]
Cigarettes per Day	8	64567670	rs1217106	A	G	0.782	34.9	3.51×10^{-9}	-0.0291	0.00493	120744	127	335003	64540485	64875487	Intergenic[YTHDF3]
Cigarettes per Day	9	136478355	rs3025343	G	A	0.117	43	5.54×10^{-11}	0.0368	0.00561	153918	8	25071	136460224	136485294	Intergenic[DBH]
Cigarettes per Day	12	4268731	rs1134691	G	A	0.057	33.1	8.74×10^{-9}	0.0505	0.00877	120744	11	126043	4156808	4282850	Intergenic[CCND2]
Cigarettes per Day	15	78806023	rs8034191	T	C	0.341	633	1.41×10^{-139}	0.0956	0.0038	153918	2	65266	78806023	78871288	Intron[AGPHD1]
Cigarettes per Day	15	78896547	rs938682	G	A	0.777	382	4.44×10^{-85}	0.0846	0.00433	153918	2	11486	78896547	78908032	Intron[CHRNA3]
Cigarettes per Day	15	78918726	rs4344703	G	T	0.68	248	7.32×10^{-56}	-0.0845	0.00536	79862	15	86006	78862845	78948850	Intron[CHRNB4]
Cigarettes per Day	17	37814687	rs3601561	G	A	0.000163	30.6	3.18×10^{-8}	1.16	0.20948	69951	8997	999750	37314891	38314640	Nonsynonymous[STARD3]
Cigarettes per Day	19	41353107	rs5611385	T	C	0.575	305	2.99×10^{-68}	0.0719	0.00412	120744	3	13822	41353107	41366928	Intron[CYP2A6]
Cigarettes per Day	19	41356379	rs2839943	A	C	0.0626	90.3	2.00×10^{-21}	-0.0799	0.0084	120744	9	39192	41333284	41372475	Upstream[CYP2A6]
Cigarettes per Day	19	41371480	rs1178244	A	G	0.0288	157	4.82×10^{-36}	-0.153	0.01217	120744	2	16948	41354533	41371480	Intergenic[CYP2A7]
Cigarettes per Day	19	41406448	rs1175404	G	A	0.0229	110	1.08×10^{-25}	-0.143	0.01361	120744	14	72606	41335269	41407874	Intergenic[CYP2A7]
Cigarettes per Day	20	61984317	rs6011779	C	T	0.806	37.9	7.43×10^{-10}	-0.0317	0.00514	120744	3	7689	61984317	61992005	Intron[CHRNA4]
Pack Years	1	110045880	rs4370783	A	C	0.115	30.7	3.06×10^{-8}	-0.0321	0.00579	146945	3	10415	110036919	110047333	Intergenic[CYB561D1]
Pack Years	2	45154418	rs7569203	A	C	0.317	30.4	3.57×10^{-8}	0.0218	0.00396	146945	18	57384	105915711	105973094	Intergenic[SIX3]
Pack Years	2	105969362	rs6215587	C	T	0.127	36.2	1.75×10^{-9}	0.0334	0.00555	146945	9	17632	45152522	45170153	Intergenic[C2orf49]
Pack Years	7	32273107	rs7806224	T	C	0.375	67.1	2.57×10^{-16}	0.0287	0.0035	174410	40	72185	32261458	32333642	Intron[PDE1C]
Pack Years	8	42550498	rs6474412	C	T	0.77	64	1.24×10^{-15}	0.0319	0.00398	177812	29	19541	42544017	42563557	Intergenic[CHRNB3]
Pack Years	9	136478355	rs3025343	G	A	0.114	90.1	2.28×10^{-21}	0.0501	0.00527	177812	1	1	136478355	136478355	Intergenic[DBH]
Pack Years	11	43603300	rs1182013	T	C	0.44	36.3	1.70×10^{-9}	-0.0224	0.00372	146945	33	126471	43601860	43728330	Intergenic[MIR670]
Pack Years	13	112179953	rs2026174	C	T	0.504	37.8	7.94×10^{-10}	0.0227	0.00369	146945	5	6331	112179953	112186283	Intergenic[C13orf16]
Pack Years	14	104146421	rs4900590	C	T	0.324	35.4	2.73×10^{-9}	0.0234	0.00394	146945	38	116843	104067895	104184737	Intron[KLC1]
Pack Years	15	78806023	rs8034191	T	C	0.338	533	5.81×10^{-118}	0.0819	0.00355	177812	8	86930	78806023	78892952	Intron[AGPHD1]
Pack Years	15	78896547	rs938682	G	A	0.773	304	5.39×10^{-68}	0.0698	0.00401	177812	2	11486	78896547	78908032	Intron[CHRNA3]
Pack Years	15	78918726	rs4344703	G	T	0.68	284	8.47×10^{-64}	-0.0667	0.00395	146945	34	329428	78642718	78972145	Intron[CHRNB4]
Pack Years	15	78962803	.	C	T	0.0511	100	1.44×10^{-23}	-0.0838	0.00838	146945	1	1	78962803	78962803	Intergenic[CHRNB4]
Pack Years	15	79004642	.	C	T	0.754	48.7	2.94×10^{-12}	0.0299	0.00428	146945	1	1	79004642	79004642	Intergenic[LOC646938]
Pack Years	15	79052227	.	T	C	0.222	29.7	5.07×10^{-8}	-0.0242	0.00444	146945	5	388223	78664005	79052227	Intron[ADAMTS7]
Pack Years	15	79078638	rs1428333	T	C	0.00431	20.9	4.77×10^{-6}	0.129	0.02816	146945	23	505793	78642718	79148510	Intron[ADAMTS7]
Pack Years	15	89928189	rs150353	T	G	0.448	30.8	2.91×10^{-8}	-0.0205	0.0037	147591	11	18402	89927560	89945961	Intron[LOC254559]
Pack Years	16	29311500	rs5367482	C	T	0.00126	31	2.60×10^{-8}	0.29	0.0521	146945	82	72034	28823097	28895130	Intron[RUNDC2C]
Pack Years	16	69972486	rs9302605	T	A	0.771	34	5.62×10^{-9}	0.0256	0.00439	146945	81	441506	69556583	69998088	Intron[WWP2]
Pack Years	16	89645437	rs369230	G	T	0.695	35	3.26×10^{-9}	0.0237	0.004	146945	1	1	89645437	89645437	Intron[CPNE7]
Pack Years	17	27447905	rs1260199	C	T	0.156	32.1	1.50×10^{-8}	0.0288	0.00509	146945	7	65845	27382061	27447905	Intron[MYO18A]
Pack Years	19	41353107	rs5611385	T	C	0.572	107	5.61×10^{-25}	0.0385	0.00373	146945	3	11879	41341229	41353107	Intron[CYP2A6]
Pack Years	19	41371480	rs1178244	A	G	0.0274	51.7	6.51×10^{-13}	-0.0813	0.01131	146945	2	16948	41354533	41371480	Intergenic[CYP2A7]
Pack Years	20	61991833	rs4549780	C	T	0.0867	68.2	1.50×10^{-16}	0.0541	0.00655	146945	3	7900	61983934	61991833	Intron[CHRNA4]
Smoking Initiation	1	43999327	rs7163660	G	T	0.255	8.67	3.24×10^{-3}	0.00772	0.00262	383631	12	44180	43955148	43999327	Intron[PTPRF]
Smoking Initiation	1	44076019	rs1121088	G	A	0.7	48.7	3.00×10^{-12}	-0.0174	0.00249	384669	9	62400	44037685	44100084	Intron[PTPRF]
Smoking Initiation	1	66470206	rs2186122	A	T	0.561	43.5	4.25×10^{-11}	0.0152	0.0023	383631	18	88784	66392405	66481188	Intron[PDE4B]
Smoking Initiation	1	73882478	rs1475064	A	G	0.615	30	4.38×10^{-8}	-0.0121	0.00221	433216	464	648514	73435171	74083684	Intergenic[LRRIQ3]
Smoking Initiation	1	74991644	rs1514175	A	G	0.575	40.4	2.07×10^{-10}	-0.0138	0.00217	433216	25	28851	74977870	75006720	Intron[FPGT-TNNI3K|TNNI3K]
Smoking Initiation	1	91191582	rs7272039	A	G	0.231	38.1	6.85×10^{-10}	-0.0167	0.00271	383631	9	6446	91189731	91196176	Intergenic[BARHL2]

Smoking Initiation	1	210359333	rs5592113	T	C	0.203	31	2.56×10^{-8}	-0.0158	0.00284	383631	16	247370	210176170	210423539	Intergenic[SYT14]
Smoking Initiation	2	615140	rs7567570	T	C	0.827	46.5	8.96×10^{-12}	0.0206	0.00302	383631	164	38442	615140	653581	Intergenic[TMEM18]
Smoking Initiation	2	45154908	rs528301	G	A	0.552	51.6	6.93×10^{-13}	0.0165	0.00229	384669	17	32700	45134871	45167570	Intergenic[SIX3]
Smoking Initiation	2	60024857	rs7585579	C	G	0.489	47.4	5.65×10^{-12}	0.0157	0.00228	383631	9	157818	60000304	60158121	Intergenic[BCL11A]
Smoking Initiation	2	60475008	rs359240	G	A	0.607	35.6	2.41×10^{-9}	0.0139	0.00234	383631	70	60815	60470926	60531740	Intergenic[BCL11A]
Smoking Initiation	2	64092819	rs1401442	C	T	0.00232	29.7	4.92×10^{-8}	-0.129	0.0237	383631	7200	999920	63592820	64592739	Intron[UGP2]
Smoking Initiation	2	104150791	rs6720941	C	T	0.461	65.4	6.06×10^{-16}	0.0185	0.00229	383631	182	241178	104057364	104298541	Intergenic[TMEM182]
Smoking Initiation	2	137542847	rs3570251	G	T	0.237	34.3	4.84×10^{-9}	0.0157	0.00269	383631	143	265371	137359506	137624876	Intergenic[THSD7B]
Smoking Initiation	2	145400317	rs1427499	A	G	0.711	40	2.57×10^{-10}	-0.0159	0.00252	383631	2	11955	145400317	145412271	Intergenic[DKFZp686O1327]
Smoking Initiation	2	146155371	rs6745444	A	G	0.482	85.5	2.36×10^{-20}	0.0211	0.00228	383631	13	13590	146143090	146156679	Intergenic[DKFZp686O1327]
Smoking Initiation	2	155742894	rs2652434	T	C	0.533	31.4	2.09×10^{-8}	0.0128	0.00229	383631	29	433295	155670875	156104169	Intergenic[KCNJ3]
Smoking Initiation	2	162891848	rs6432708	C	T	0.581	49.1	2.38×10^{-12}	-0.0162	0.00231	383631	30	93268	162798581	162891848	Intron[DPP4]
Smoking Initiation	3	5723818	rs4479577	C	T	0.482	38.7	4.85×10^{-10}	0.0142	0.00228	383631	12	6331	5720948	5727278	Intergenic[EDEM1]
Smoking Initiation	3	25193102	rs904592	C	T	0.482	30.2	3.80×10^{-8}	0.0126	0.00228	383631	109	147649	25147940	25295588	Intergenic[RARB]
Smoking Initiation	3	52874288	rs6445538	T	C	0.233	42.4	7.58×10^{-11}	0.0165	0.00254	433216	9	39219	52852538	52891756	Utr3[TMEM110]
Smoking Initiation	3	61819535	rs3599516	A	G	0.402	33.3	8.00×10^{-9}	0.0134	0.00233	383631	57	69201	61809160	61878360	Intron[PTPRG]
Smoking Initiation	3	85460131	rs1549979	C	T	0.622	71.3	3.12×10^{-17}	-0.0187	0.00222	433216	307	242238	85407980	85650217	Intron[CADM2]
Smoking Initiation	3	85985324	rs6668080	G	T	0.399	35.8	2.21×10^{-9}	-0.0139	0.00233	383631	85	218344	85868529	86086872	Intron[CADM2]
Smoking Initiation	3	117639575	rs6226476	G	A	0.149	37.5	9.05×10^{-10}	-0.0196	0.0032	383631	155	449454	117517339	117966792	Intergenic[IGSF11]
Smoking Initiation	3	118302515	rs1205387	T	G	0.545	33.3	7.95×10^{-9}	0.0132	0.00229	383631	59	107857	118294235	118402091	Intergenic[IGSF11]
Smoking Initiation	3	157393770	rs963354	C	A	0.672	33.5	6.98×10^{-9}	0.0141	0.00243	384669	37	200909	157377765	157578673	Intergenic[C3orf55]
Smoking Initiation	4	28540930	rs1312532	A	C	0.258	31.9	1.63×10^{-8}	0.0147	0.00261	383631	183	544291	28191284	28735574	Intergenic[MIR4275]
Smoking Initiation	4	57740334	rs6835108	A	G	0.215	36	1.94×10^{-9}	-0.0167	0.00278	383631	16	5239	57738345	57743583	Intergenic[REST]
Smoking Initiation	4	143617304	rs1194339	T	C	0.628	35.8	2.22×10^{-9}	0.0141	0.00236	383631	19	122574	143500223	143622796	Intron[INPP4B]
Smoking Initiation	4	147948150	rs3827592	G	A	0.351	40.1	2.46×10^{-10}	-0.0151	0.00239	383631	89	219281	147748826	147968106	Intergenic[TTC29]
Smoking Initiation	5	22219503	rs4518351	A	G	0.444	32.4	1.28×10^{-8}	0.0131	0.00229	384669	21	71936	22160029	22231964	Intron[CDH12]
Smoking Initiation	5	60121470	rs6237207	A	T	0.148	34.7	3.82×10^{-9}	0.019	0.00322	383631	71	421607	60057718	60479324	Intron[ELOVL7]
Smoking Initiation	5	80263403	rs1335701	G	A	0.635	30.4	3.54×10^{-8}	0.0131	0.00237	383631	6	23327	80240539	80263865	Intron[RASGRF2]
Smoking Initiation	5	87781168	rs1004461	C	T	0.432	47.5	5.39×10^{-12}	-0.0159	0.0023	383631	21	100175	87680994	87781168	Intergenic[LOC645323]
Smoking Initiation	5	94202167	rs27003	T	C	0.694	33.3	8.02×10^{-9}	0.0143	0.00248	383631	4	27588	94198290	94225877	Intron[MCTP1]
Smoking Initiation	5	106825618	rs7278962	T	A	0.138	35.7	2.33×10^{-9}	-0.0198	0.00331	383631	7	15162	106825618	106840779	Intron[EFNA5]
Smoking Initiation	5	157743295	rs1113503	C	T	0.306	37.7	8.17×10^{-10}	0.0152	0.00248	383631	29	20603	157722715	157743317	Intergenic[EBF1]
Smoking Initiation	5	166996722	rs1549212	C	T	0.626	37.1	1.14×10^{-9}	-0.0144	0.00236	383631	22	9801	166988585	166998385	Intron[ODZ2]
Smoking Initiation	6	28366151	rs2232423	A	G	0.111	32.1	1.43×10^{-8}	-0.0194	0.00342	433216	502	998466	27866384	28864849	Nonsynonymous[ZSCAN12]
Smoking Initiation	6	29548089	rs926552	G	A	0.134	32.8	1.00×10^{-8}	-0.0181	0.00315	433216	465	981009	29057639	30038647	Intergenic[SNORD32B]
Smoking Initiation	6	30857894	rs1264322	G	A	0.146	31.7	1.77×10^{-8}	-0.0171	0.00304	433216	55	763863	30365448	31129310	Intron[DDR1]
Smoking Initiation	6	31686497	rs15574	G	A	0.194	33.5	7.12×10^{-9}	0.0159	0.00275	422914	18	540260	31605448	32145707	Utr3[LY6G6C]
Smoking Initiation	6	48121843	NA	T	TA	0.1	30.5	3.26×10^{-8}	0.0388	0.00701	112811	4	110489	48121843	48232331	Intergenic[C6orf138]
Smoking Initiation	6	67546542	rs1220553	T	C	0.387	36.4	1.60×10^{-9}	0.0141	0.00234	383631	105	150542	67405337	67555878	Intergenic[MCART3P]
Smoking Initiation	6	93833436	rs6938042	T	G	0.339	36	2.01×10^{-9}	0.0144	0.00241	384669	3	9011	93833436	93842446	Intergenic[EPHA7]
Smoking Initiation	6	98888740	rs2132029	C	T	0.318	39.5	3.20×10^{-10}	0.0154	0.00245	383631	103	261293	98631890	98893182	Intergenic[POU3F2]
Smoking Initiation	6	101153907	rs6922219	G	C	0.47	31.2	2.28×10^{-8}	0.0128	0.00229	383631	183	385831	100953148	101338978	Intron[ASCC3]
Smoking Initiation	6	111644332	rs240963	T	C	0.841	72.1	2.03×10^{-17}	-0.0265	0.00312	383631	37	287818	111578490	111866307	Intron[REV3L]
Smoking Initiation	7	1701592	rs1025500	C	T	0.432	42.2	8.28×10^{-11}	-0.015	0.0023	383631	46	45754	1662327	1708080	Intergenic[TFAMP1]
Smoking Initiation	7	3448973	rs2056475	A	G	0.746	42.9	5.65×10^{-11}	-0.0172	0.00262	383631	143	245471	3314356	3559826	Intron[SDK1]
Smoking Initiation	7	96624257	rs2240294	T	A	0.444	30	4.25×10^{-8}	-0.0126	0.0023	383631	8	16261	96622007	96638267	Intron[DLX6-AS1]
Smoking Initiation	7	117523709	rs1023301	A	G	0.504	43	5.53×10^{-11}	0.015	0.00228	383631	30	69766	117497811	117567576	Intergenic[CTTNBP2]
Smoking Initiation	8	10153082	rs1715163	C	T	0.281	35.9	2.12×10^{-9}	-0.0152	0.00254	383631	48	474267	9803712	10277978	Intron[MSRA]
Smoking Initiation	8	10836472	rs7010246	G	A	0.365	30.7	3.04×10^{-8}	-0.0131	0.00237	383631	182	660014	10519445	11179458	Intron[XKR6]
Smoking Initiation	8	27426077	rs1565735	T	A	0.201	67.8	1.77×10^{-16}	-0.0235	0.00285	383631	2	16051	27426077	27442127	Intergenic[EPHX2]
Smoking Initiation	8	59817068	rs1562612	A	G	0.504	37.4	9.73×10^{-10}	-0.0139	0.00228	384669	30	18107	59799367	59817473	Intron[TOX]

Smoking Initiation	8	92775372	rs1095680	T	G	0.422	36.1	1.86×10^{-9}	-0.0131	0.00218	433216	124	313254	92464180	92777433	Intergenic[RUNX1T1]
Smoking Initiation	8	93201036	rs1899896	C	T	0.297	34.9	3.41×10^{-9}	0.0148	0.0025	383631	37	77665	93184065	93261729	Intergenic[RUNX1T1]
Smoking Initiation	9	16758107	rs1096256	G	C	0.149	38.7	5.00×10^{-10}	0.0199	0.00321	383631	4	10708	16749265	16759972	Intron[BNC2]
Smoking Initiation	9	86761781	rs1246264	T	C	0.695	38.8	4.82×10^{-10}	0.0154	0.00248	384669	9	33583	86728457	86762039	Intergenic[SLC28A3]
Smoking Initiation	9	128134034	rs7870475	T	C	0.475	44.4	2.73×10^{-11}	0.0152	0.00229	383631	86	126016	128011403	128137418	Intergenic[GAPVD1]
Smoking Initiation	9	137975748	rs7019640	C	T	0.23	32.7	1.10×10^{-8}	0.0155	0.00271	383631	8	7512	137970849	137978360	Intron[OLFM1]
Smoking Initiation	10	21783634	rs1277022	G	A	0.315	42.2	8.24×10^{-11}	0.016	0.00246	383631	60	403745	21746967	22150711	Utr3[C10orf114]
Smoking Initiation	10	27397454	NA	C	T	0.093	31.1	2.46×10^{-8}	0.0404	0.00725	112811	3	42396	27397454	27439849	Intergenic[YME1L1]
Smoking Initiation	10	63674885	rs7921378	G	C	0.481	42.8	6.06×10^{-11}	-0.0149	0.00228	383631	16	14732	63666691	63681422	Intron[ARID5B]
Smoking Initiation	10	104045951	NA	GT	G	0.828	31.7	1.75×10^{-8}	-0.0314	0.00557	112811	2	83214	104413385	104496598	Deletion[GBF1]
Smoking Initiation	10	104643241	rs1278378	G	A	0.257	59.9	1.00×10^{-14}	0.0202	0.00261	383631	55	349356	104611764	104961119	Intron[AS3MT|C10orf32-AS3I
Smoking Initiation	10	125680419	rs9423279	C	G	0.657	33.3	7.90×10^{-9}	-0.0139	0.0024	383631	4	20524	125680419	125700942	Intergenic[CPXM2]
Smoking Initiation	11	7950797	rs4523689	A	G	0.393	32.4	1.25×10^{-8}	-0.0133	0.00234	383631	20	314107	7940327	8254433	Intergenic[OR10A6]
Smoking Initiation	11	20129311	rs3589196	G	A	0.0719	32.4	1.27×10^{-8}	-0.0237	0.00416	433216	1	1	20129311	20129311	Nonsynonymous[NAV2]
Smoking Initiation	11	27694241	rs2049045	G	C	0.187	46.4	9.77×10^{-12}	-0.0199	0.00293	383631	25	98804	27610041	27708844	Intron[BDNF|BDNF-AS1]
Smoking Initiation	11	85927213	rs1466802	T	C	0.775	38.1	6.68×10^{-10}	0.0169	0.00274	383631	21	398675	85927213	86325887	Intergenic[EED]
Smoking Initiation	11	112839532	rs7948789	A	G	0.386	116	6.11×10^{-27}	0.0252	0.00234	383631	81	77701	112835024	112912724	Intron[NCAM1]
Smoking Initiation	12	56412487	rs1701704	T	G	0.341	30.9	2.70×10^{-8}	-0.0126	0.00227	433216	27	112270	56369911	56482180	Intergenic[IKZF4]
Smoking Initiation	12	57976118	rs1117225	A	G	0.257	29.8	4.80×10^{-8}	0.0143	0.00261	383631	14	32524	57975155	58007678	Intron[KIF5A]
Smoking Initiation	12	69686203	rs317660	G	A	0.718	35.8	2.23×10^{-9}	-0.0152	0.00254	383631	25	77611	69637847	69715457	Intergenic[CPSF6]
Smoking Initiation	12	111904371	rs4766578	T	A	0.504	30.8	2.83×10^{-8}	-0.0127	0.00228	384669	14	704633	111495518	112200150	Intron[ATXN2]
Smoking Initiation	12	121111214	rs3406737	C	T	0.14	30.8	2.80×10^{-8}	0.0183	0.00329	383631	20	298946	121099149	121398094	Intergenic[CABP1]
Smoking Initiation	12	133451277	rs8003224	G	A	0.137	30.4	3.54×10^{-8}	0.0183	0.00332	383631	12	117588	133362170	133479757	Intron[CHFR]
Smoking Initiation	13	38359684	rs4460944	T	C	0.527	31.8	1.68×10^{-8}	0.0129	0.00229	383631	48	11493	38349236	38360728	Intron[TRPC4]
Smoking Initiation	13	100548329	rs7322872	C	T	0.791	34	5.55×10^{-9}	-0.0164	0.00281	383631	6	5127	100545308	100550434	Utr3[CLYBL]
Smoking Initiation	13	101179012	rs837333	T	C	0.477	35.7	2.24×10^{-9}	0.0137	0.00229	383631	56	108502	101150926	101259427	Intron[PCCA]
Smoking Initiation	14	29500106	rs7141236	T	C	0.187	34.6	4.01×10^{-9}	-0.0172	0.00292	384669	72	219248	29316842	29536089	Intergenic[C14orf23]
Smoking Initiation	14	90030280	rs171460	G	A	0.26	31	2.52×10^{-8}	-0.0145	0.0026	383631	13	6221	90027413	90033633	Intron[FOXN3]
Smoking Initiation	15	47616380	rs6493272	T	C	0.388	53.4	2.73×10^{-13}	0.0171	0.00234	383631	86	356232	47615220	47971451	Intron[SEMA6D]
Smoking Initiation	15	47935843	rs1435741	G	A	0.43	55	1.18×10^{-13}	0.0171	0.00231	383631	86	356232	47615220	47971451	Intron[SEMA6D]
Smoking Initiation	15	54119670	rs766614	T	G	0.212	31.8	1.75×10^{-8}	0.0157	0.00279	384669	6	436576	53775670	54212245	Intergenic[WDR72]
Smoking Initiation	16	735921	rs763053	T	C	0.225	34.3	4.70×10^{-9}	-0.016	0.00273	384669	33	54884	694174	749057	Synonymous[WDR24]
Smoking Initiation	16	18058548	rs7787847	T	A	0.0858	39.5	3.31×10^{-10}	-0.0256	0.00408	383631	1	1	18058548	18058548	Intergenic[NOMO2]
Smoking Initiation	16	69567781	rs1164657	G	A	0.439	39.6	3.13×10^{-10}	0.0145	0.0023	383631	32	196550	69567411	69763960	Intergenic[NFAT5]
Smoking Initiation	18	77458644	rs522180	C	T	0.353	45.6	1.47×10^{-11}	0.0161	0.00239	383631	18	47828	77454237	77502064	Intron[CTDP1]
Smoking Initiation	19	4474725	rs7660858	C	A	0.0482	32.6	1.15×10^{-8}	-0.0304	0.00533	383631	1	1	4474725	4474725	Intron[HDGFRP2]
Smoking Initiation	20	29893501	rs6061162	C	T	0.912	29.8	4.83×10^{-8}	0.022	0.00402	383631	27	498697	29471217	29969913	Intron[DEFB116]
Smoking Initiation	20	31140165	rs4911241	C	T	0.239	51.9	5.88×10^{-13}	0.0182	0.00252	433216	13	76764	31093514	31170277	Intergenic[LOC149950]
Drinks per Week	1	4569436	rs780569	T	A	0.708	33.2	8.47×10^{-9}	-0.0161	0.00279	311126	25	171838	4548453	4720290	Intergenic[LOC284661]
Drinks per Week	1	95730588	rs3886141	G	C	0.409	31.5	1.96×10^{-8}	-0.0145	0.00258	311126	25	116820	95622341	95739160	Intergenic[RWDD3]
Drinks per Week	1	165119792	rs1075366	G	A	0.684	33.2	8.36×10^{-9}	-0.0157	0.00273	311126	5	5947	165117049	165122995	Intergenic[LMX1A]
Drinks per Week	2	27730940	rs1260326	T	C	0.608	176	4.58×10^{-40}	0.0321	0.00242	357854	1	1	27730940	27730940	Nonsynonymous[GCKR]
Drinks per Week	2	44271496	rs7512054	C	T	0.0302	33.7	6.29×10^{-9}	-0.043	0.00741	311126	28	232660	44067220	44299879	Intergenic[LRPPRC]
Drinks per Week	2	45154908	rs528301	G	A	0.553	69.6	7.12×10^{-17}	0.0213	0.00255	311789	5	17874	45141218	45159091	Intergenic[SIX3]
Drinks per Week	2	144272376	rs4233567	C	T	0.364	37.5	8.92×10^{-10}	-0.0161	0.00263	311126	18	109768	144162609	144272376	Intron[ARHGAP15]
Drinks per Week	3	85049088	rs1872552	G	A	0.399	44.4	2.73×10^{-11}	0.0172	0.00259	311789	165	562933	84986088	85549020	Intron[CADM2]
Drinks per Week	3	85593584	rs6780346	C	T	0.621	41.1	1.47×10^{-10}	-0.0168	0.00261	311126	417	247245	85403892	85651136	Intron[CADM2]
Drinks per Week	4	39413780	rs2871282	G	A	0.608	187	1.67×10^{-42}	0.0355	0.0026	311126	2	1214	39413780	39414993	Intron[KLB]
Drinks per Week	4	42117559	rs1685402	G	A	0.13	37.3	1.02×10^{-9}	0.023	0.00377	311126	6	58502	42117559	42176060	Utr3[BEND4]
Drinks per Week	4	99678179	rs1170532	C	T	0.00182	0.375	5.40×10^{-1}	0.0182	0.02972	311126	34	499164	99678179	100177342	Intergenic[TSPAN5]
Drinks per Week	4	100239319	rs1229984	T	C	0.98	788.2	2.27×10^{-173}	0.246	0.00878	334588	1	1	100239319	100239319	Nonsynonymous[ADH1B]

Drinks per Week	4	100291643	rs6836438	G	T	0.134	51.8	6.19×10^{-13}	-0.0268	0.00372	311126	2	37124	100254520	100291643	Intergenic[ADH1C]
Drinks per Week	4	100291659	rs5602200	T	C	0.113	4.09	4.30×10^{-2}	0.00809	0.004	311126	6	76073	100215587	100291659	Intergenic[ADH1C]
Drinks per Week	4	100407531	rs7826106	C	T	0.00116	10.4	1.28×10^{-3}	-0.12	0.03722	311126	1	1	100390796	100390796	Intergenic[C4orf17]
Drinks per Week	4	101347141	rs1193951	G	A	0.0975	7.64	5.71×10^{-3}	0.0118	0.00427	311126	46	147760	101288442	101436201	Intron[EMCN]
Drinks per Week	4	102702364	rs1310568	T	G	0.0591	35.4	2.72×10^{-9}	-0.032	0.00538	311126	3	196434	103001649	103198082	Intergenic[BANK1]
Drinks per Week	4	143617304	rs1194339	T	C	0.629	32.1	1.43×10^{-8}	0.0149	0.00262	311126	136	231745	143487451	143719195	Intron[INPP4B]
Drinks per Week	6	12903957	rs9349379	A	G	0.406	30.8	2.85×10^{-8}	0.0134	0.00241	357854	5	108927	12795031	12903957	Intron[PHACTR1]
Drinks per Week	7	69793380	rs6978841	T	A	0.184	31.3	2.24×10^{-8}	-0.0183	0.00327	311126	77	382043	69594503	69976545	Intron[AUTS2]
Drinks per Week	7	153489725	rs6969458	G	A	0.466	44.6	2.45×10^{-11}	0.017	0.00254	311126	31	31145	153463438	153494582	Intergenic[DPP6]
Drinks per Week	9	108755622	rs5593221	A	G	0.746	32.2	1.41×10^{-8}	0.0165	0.00291	311126	11	48497	108707126	108755622	Intergenic[TMEM38B]
Drinks per Week	9	109330236	rs4743009	G	A	0.201	33.5	6.98×10^{-9}	-0.0183	0.00317	311126	38	93411	109321180	109414590	Intergenic[ZNF462]
Drinks per Week	10	102626510	rs6187351	G	T	0.326	33.5	7.14×10^{-9}	-0.0157	0.0027	311126	3	8966	102626510	102635475	Intergenic[PAX2]
Drinks per Week	11	33204159	rs4756121	G	C	0.688	42.8	6.08×10^{-11}	0.0179	0.00274	311126	83	104350	33113110	33217459	Intron[LOC338739]
Drinks per Week	11	47457539	rs1160468	A	G	0.32	40.8	1.69×10^{-10}	-0.0174	0.00272	311126	126	543224	47375193	47918416	Intergenic[RAPSN]
Drinks per Week	11	113412443	rs4309187	A	C	0.684	45.9	1.25×10^{-11}	0.0185	0.00273	311126	21	106298	113317745	113424042	Intergenic[DRD2]
Drinks per Week	11	113660576	rs1713676	A	G	0.521	33.7	6.31×10^{-9}	-0.0147	0.00254	311789	184	305363	113456292	113761654	Intergenic[USP28]
Drinks per Week	11	116073667	rs1894071	A	G	0.857	30.2	3.94×10^{-8}	0.0199	0.00362	311126	8	28722	116073667	116102388	Intergenic[LOC283143]
Drinks per Week	11	121501406	rs1079044	T	C	0.465	33.6	6.91×10^{-9}	0.0147	0.00254	311126	28	58174	121495141	121553314	Utr3[SORL1]
Drinks per Week	14	57281154	rs962961	C	T	0.326	34.7	3.88×10^{-9}	-0.0159	0.00271	311126	8	22904	57269825	57292728	Intron[OTX2OS1]
Drinks per Week	14	58833909	rs5704421	G	A	0.284	35.1	3.17×10^{-9}	0.0166	0.00281	311126	67	165426	58668484	58833909	Intron[ARID4A]
Drinks per Week	14	94844947	rs2892947	C	T	0.0198	44.8	2.22×10^{-11}	-0.0568	0.00849	357854	1	1	94844947	94844947	Nonsynonymous[SERPINA1]
Drinks per Week	16	24810681	rs1717707	C	T	0.0554	34.9	3.45×10^{-9}	-0.0327	0.00554	311126	3	117634	24693048	24810681	Intron[TNRC6A]
Drinks per Week	16	28539848	rs4788084	C	T	0.418	31.1	2.46×10^{-8}	-0.0134	0.0024	355949	145	582460	28336882	28919341	Intergenic[NUPR1]
Drinks per Week	16	69745145	rs1800566	G	A	0.183	31.5	1.99×10^{-8}	0.0172	0.00306	357854	103	237972	69549187	69787158	Nonsynonymous[NQO1]
Drinks per Week	16	72356964	rs1164857	T	C	0.107	31.5	2.04×10^{-8}	0.023	0.00409	311126	18	626565	72067522	72694086	Intergenic[PMFBP1]
Drinks per Week	16	73912588	rs1104608	G	C	0.426	40.6	1.87×10^{-10}	-0.0163	0.00256	311126	3	17543	73895046	73912588	Intergenic[PSMD7]
Drinks per Week	17	43754850	rs6205569	T	C	0.225	99.7	1.78×10^{-23}	-0.0303	0.00303	311126	1459	671979	43572419	44244397	Intron[CRHR1]
Drinks per Week	17	44285531	rs5597542	G	A	0.225	94.7	2.27×10^{-22}	-0.0295	0.00303	311789	1300	548426	43785627	44334052	Intron[KIAA1267]
Drinks per Week	17	44843136	rs199528	C	T	0.223	88.4	5.27×10^{-21}	-0.0286	0.00304	311126	47	517285	44345063	44862347	Intron[WNT3]
Drinks per Week	18	40735531	rs6209206	G	A	0.373	31.9	1.60×10^{-8}	0.0148	0.00262	311126	80	47623	40702833	40750455	Intergenic[RIT2]
Drinks per Week	19	49248730	rs838145	G	A	0.541	46.6	8.86×10^{-12}	-0.0174	0.00254	311126	16	57061	49206417	49263477	Intron[IZUMO1]
Drinks per Week	20	18693080	rs6221702	A	G	0.396	32.6	1.10×10^{-8}	-0.0148	0.00259	311126	102	147966	18569599	18717564	Intron[DTD1]
Drinks per Week	20	31607551	rs6173434	G	C	0.0598	32.7	1.09×10^{-8}	-0.0285	0.00499	357854	6	373090	31248265	31621354	Nonsynonymous[BPIL1]
Drinks per Week	22	41941243	rs9607812	G	A	0.187	38	7.21×10^{-10}	-0.02	0.00325	311126	147	222403	41749630	41972032	Intergenic[POLR3H]
Drinks per Week	22	46481623	rs1788469	G	A	0.248	32.7	1.08×10^{-8}	-0.0168	0.00294	311126	2	1022	46480602	46481623	Intergenic[LOC400931]

Supplemental Table 7. Bayesian Fine Mapping with Prior Informed by Functional Annotations, all Variants with Posterior Probability of Association Greater than .80 are shown

PHENOTYPE	POSITION	RS Number	REF/ALT	AF	P-VALUE	BETA	SE	Posterior Probablity of Association	Annotation
Cigarettes per Day	8:42550498	rs6474412	C/T	0.775	3.40×10^{-16}	0.035	0.0043	0.95	Intergenic[Inte
Cigarettes per Day	9:136478355	rs3025343	G/A	0.117	5.50×10^{-11}	0.037	0.0056	0.92	Intergenic[Inte
Cigarettes per Day	15:78882925	rs16969968	G/A	0.337	2.50×10^{-139}	0.096	0.0038	0.92	Asp398Asn[CH
Cigarettes per Day	15:79211424	rs147125452	T/C	0.026	8.50×10^{-10}	-0.078	0.0127	0.86	Intergenic[Inte
Cigarettes per Day	19:41353107	rs56113850	T/C	0.575	3.00×10^{-68}	0.072	0.0041	1	Intron[CYP2A6]
Cigarettes per Day	20:61984317	rs6011779	C/T	0.806	7.40×10^{-10}	-0.032	0.0051	0.83	Intron[CHRNA4
Pack Years	9:136478355	rs3025343	G/A	0.114	2.30×10^{-21}	0.05	0.0053	0.98	Intergenic[Inte
Pack Years	15:78806023	rs8034191	T/C	0.338	5.80×10^{-118}	0.082	0.0036	0.99	Intron[AGPHD1
Pack Years	19:41353107	rs56113850	T/C	0.572	5.60×10^{-25}	0.038	0.0037	0.8	Intron[CYP2A6]
Pack Years	20:61991833	rs45497800	C/T	0.087	1.50×10^{-16}	0.054	0.0066	1	Intron[CHRNA4
Smoking Initiation	1:94078440	rs236345	A/G	0.221	9.10×10^{-8}	-0.015	0.0027	0.84	Intron[BCAR3]
Smoking Initiation	1:108507209	rs56193625	G/A	0.0556	3.80×10^{-7}	-0.025	0.005	0.85	Intron[VAV3]
Smoking Initiation	1:179783167	rs147052174	G/T	0.0181	1.20×10^{-7}	0.043	0.008	1	Gly116Val[FAM
Smoking Initiation	1:237841390	rs34967813	A/G	0.312	8.10×10^{-7}	-0.011	0.0023	0.98	Gln2958Arg[RY
Smoking Initiation	2:64092819	rs140144265	C/T	0.00232	4.90×10^{-8}	-0.129	0.0237	0.91	Intron[UGP2]
Smoking Initiation	5:106834363	rs72789632	C/T	0.13	2.40×10^{-9}	-0.02	0.0034	0.93	Intron[EFNA5]
Smoking Initiation	5:118691561	rs139564712	C/T	0.00055	2.40×10^{-7}	0.464	0.0897	0.93	Intron[TNFAIP8
Smoking Initiation	7:121983843	rs2289705	A/G	0.214	5.50×10^{-8}	0.015	0.0028	0.86	Intron[CADPS2
Smoking Initiation	8:27442127	rs73229090	C/A	0.118	3.40×10^{-15}	-0.028	0.0035	0.85	Intergenic[Inte
Smoking Initiation	9:16749265	rs56348592	A/G	0.145	5.80×10^{-10}	0.02	0.0032	0.8	Intron[BNC2]
Smoking Initiation	9:86755967	rs1030856	A/G	0.68	2.40×10^{-9}	0.015	0.0024	0.85	Intergenic[Inte
Smoking Initiation	11:20129311	rs35891966	G/A	0.0719	1.30×10^{-8}	-0.024	0.0042	1	Val2251Ile[NAV
Smoking Initiation	11:27679916	rs6265	C/T	0.19	1.90×10^{-10}	-0.017	0.0027	0.83	Val66Met[BDN
Smoking Initiation	14:31774324	rs61754158	C/T	0.009	1.40×10^{-6}	-0.055	0.0114	0.87	Gly1676Arg[HE
Smoking Initiation	19:4474725	rs76608582	C/A	0.0482	1.20×10^{-8}	-0.03	0.0053	0.95	Intron[HDGFRP
Smoking Initiation	19:41084693	rs814518	G/A	0.304	5.50×10^{-8}	-0.013	0.0025	0.84	Intron[SHKBP1]
Drinks per Week	2:27730940	rs1260326	T/C	0.61	4.60×10^{-40}	0.032	0.0024	1	Leu446Pro[GCK
Drinks per Week	4:39413780	rs28712821	G/A	0.608	1.70×10^{-42}	0.035	0.0026	0.94	Intron[KLB]
Drinks per Week	4:42117559	rs16854020	G/A	0.13	1.00×10^{-9}	0.023	0.0038	0.98	Utr3[BEND4]
Drinks per Week	4:100239319	rs1229984	T/C	0.98	2.30×10^{-173}	0.246	0.0088	1	His48Arg[ADH1
Drinks per Week	4:103198082	rs13135092	A/G	0.082	3.70×10^{-16}	-0.038	0.0046	0.91	Intron[SLC39A8
Drinks per Week	7:20824614	rs34908430	C/T	0.299	5.60×10^{-8}	-0.015	0.0027	0.99	Synonymous[SI
Drinks per Week	7:99229435	rs3735453	T/C	0.077	4.30×10^{-6}	-0.022	0.0048	0.84	Utr3[ZNF498]
Drinks per Week	12:50263148	rs7132908	G/A	0.375	1.90×10^{-7}	-0.013	0.0025	0.89	Utr3[FAIM2]
Drinks per Week	14:94844947	rs28929474	C/T	0.02	2.20×10^{-11}	-0.057	0.0085	1	Glu366Lys[SERI
Drinks per Week	16:69745145	rs1800566	G/A	0.183	2.00×10^{-8}	0.017	0.0031	0.97	Pro153Ser[NQO
Drinks per Week	19:49248730	rs838145	G/A	0.541	8.90×10^{-12}	-0.017	0.0025	0.95	Intron[IZUMO1
Drinks per Week	20:25096789	rs1884265	A/C	0.79	9.50×10^{-8}	0.017	0.0031	0.87	Intergenic[Inte